EXHIBIT 1

The United States of America

The Director of the United States Patent and Trademark Office

Has received an application for a patent for a new and useful invention. The title and description of the invention are enclosed. The requirements of law have been complied with, and it has been determined that a patent on the invention shall be granted under the law.

Therefore, this

United States Patent

Grants to the person(s) having title to this patent the right to exclude others from making, using, offering for sale, or selling the invention throughout the United States of America or importing the invention into the United States of America for the term set forth below, subject to the payment of maintenance fees as provided by law.

If this application was filed prior to June 8, 1995, the term of this patent is the longer of seventeen years from the date of grant of this patent or twenty years from the earliest effective U.S. filing date of the application, subject to any statutory extension.

If this application was filed on or after June 8, 1995, the term of this patent is twenty years from the U.S. filing date, subject to any statutory extension. If the application contains a specific reference to an earlier filed application or applications under 35 U.S.C. 120, 121 or 365(c), the term of the patent is twenty years from the date on which the earliest application was filed, subject to any statutory extensions.

Director of the United States Patent and Trademark Office

SS 00001

US006501966B1

(12) **United States Patent**
Bareis et al.

(10) **Patent No.: US 6,501,966 B1**
(45) **Date of Patent: Dec. 31, 2002**

(54) **SPEECH RECOGNITION SYSTEM FOR ELECTRONIC SWITCHES IN A NON-WIRELINE COMMUNICATIONS NETWORK**

(75) Inventors: **Bernard F. Bareis**, Dallas, TX (US); **Peter J. Foster**, Dallas, TX (US); **Thomas B. Schalk**, Dallas, TX (US)

(73) Assignee: **Koninklijke Philips Electronics N.V.**, Eindhoven (NL)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 84 days.

(21) Appl. No.: **09/722,810**

(22) Filed: **Nov. 27, 2000**

Related U.S. Application Data

(63) Continuation of application No. 08/914,440, filed on Aug. 19, 1997, now Pat. No. 6,157,848, which is a continuation of application No. 08/216,009, filed on Mar. 22, 1994, now Pat. No. 5,659,597, which is a continuation of application No. 07/867,873, filed on Apr. 13, 1992, now Pat. No. 5,297,183.

(51) **Int. Cl.**[7] **H04B 1/38**; H04M 1/00

(52) **U.S. Cl.** **455/563**; 455/414; 379/88.01; 379/88.03

(58) **Field of Search** 455/414, 418, 455/563; 379/88.01, 88.02, 88.03, 88.04

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,673,331 A		6/1972	Hair et al.	179/15 B
3,928,724 A		12/1975	Byram et al.	179/1 SD
4,385,359 A		5/1983	Watari et al.	364/513
4,587,670 A		5/1986	Levinson et al.	381/43
4,827,500 A		5/1989	Binkerd et al.	379/88
4,922,538 A		5/1990	Tchorzewski	381/42
4,949,374 A	*	8/1990	Ishii et al.	455/563
4,961,212 A	*	10/1990	Marui et al.	379/354
5,054,053 A	*	10/1991	Sakanishi et al.	379/354
5,148,471 A	*	9/1992	Metroka et al.	379/88.02

(List continued on next page.)

OTHER PUBLICATIONS

"Network Protocols for the Cellular Packet Switch", by Meier–Hellstern et al, EEE Transactions on Communications, vol. 42, No. 2/3/4, Feb./Mar./Apr. 1994, pp. 1235–1243.

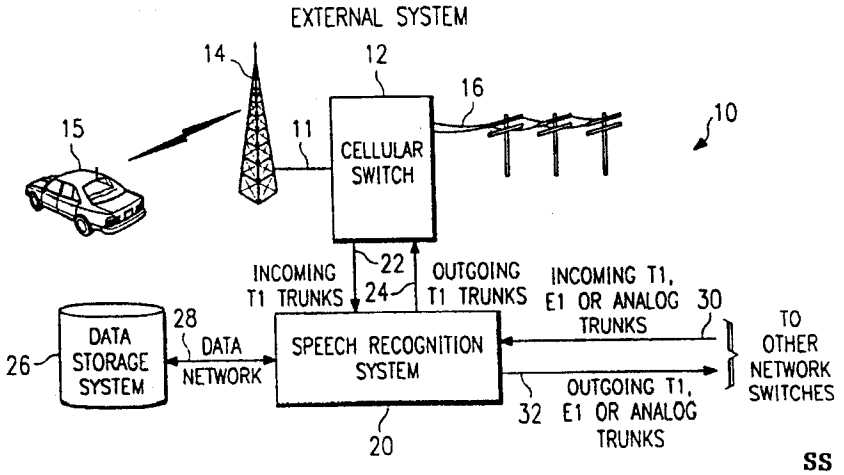
(List continued on next page.)

Primary Examiner—Lee Nguyen
Assistant Examiner—Simon Nguyen

(57) **ABSTRACT**

An advanced telecommunications system is provided for the recognizing of spoken commands over a cellular telephone, satellite telephone, or personal communications network. In the cellular application, for example, a Speech Recognition System interconnects either internally with or as an external peripheral to a cellular telecommunications switch. The Speech Recognition System includes an administrative subsystem, a call processing subsystem, a speaker-dependent recognition subsystem, a speaker-independent recognition subsystem, and a data storage subsystem. The Speech Recognition System also allows for increased efficiency in the cellular telephone network by integrating with the switch or switches as a shared resource. The administrative subsystem of the Speech Recognition System is used to keep statistical logs of pertinent call information. Pre-recorded instructional messages are stored in the memory of the call processing subsystem for instructing a user on his or her progress in using the system. The speaker-independent recognition subsystem allows the user to interact with the system employing non-user specific functions. User specific functions are controlled with the speaker-dependent recognition subsystem. User specific attributes collected by the recognition subsystems are stored in the data storage subsystem.

30 Claims, 12 Drawing Sheets



SS 00002

U.S. PATENT DOCUMENTS

5,182,765	A	*	1/1993	Ishii et al.	379/354
5,195,090	A		3/1993	Bolliger et al.	370/94
5,199,062	A		3/1993	Von Meister et al.	379/67
5,222,121	A	*	6/1993	Shimada	379/354
5,283,833	A		2/1994	Church et al.	381/41
5,289,530	A		2/1994	Reese	379/88
5,297,183	A	*	3/1994	Bareis et al.	379/88.02
5,301,223	A		4/1994	Amadon et al.	379/58
5,335,313	A	*	8/1994	Douglas	379/355.01
5,365,574	A	*	11/1994	Hunt et al.	379/189
5,371,901	A		12/1994	Reed et al.	455/69
5,390,278	A	*	2/1995	Gupta et al.	379/88.01
5,659,597	A	*	8/1997	Bareis et al.	379/88.01
6,157,848	A	*	12/2000	Bareis et al.	379/88.01

OTHER PUBLICATIONS

"The 5ESS® Wireless Mobile Switching Center", by Gauldin et al, AT&T Technical Journal, Jul./Aug. 1993, pp. 38–47.

"Environment Adaptation for Speech Recognition in Noise", by Blanchet et al, Signal Processing VI: Theories and Applications, 1992, pp. 391–394.

"HMM Modeling for Speaker Independent Voice Dialing in Car Environment", by Fissore et al, IEEE 1992, pp. I–249–I–252.

"Hands–Free Voice Communication in an Automobile with a Microphone Array", by Oh et al, IEEE 1992, pp. I–281–I–I–284.

"A Multi–DSP Implementation of a Broad–Band Adaptive Beamformer for Use in a Hands–Free Mobile Radio Telephone", by Claesson et al, IEEE Transactions on Vehicular Technology, vol. 40, No. 1, Feb. 1991, pp. 194–202.

"Automating Services with Speech Recognition over the Public Switched Telephone Network: Human Factors Considerations", by Karis et al, IEEE Journal on Selected Areas in Communications, vol. 9, No. 4, May 1991, pp. 574–585.

"Freedom Doesn't Come Easy", by Shandle, Electronics, Mar. 1991, pp. 45–48.

"Wireless Network Directions" by Ross, IEEE Communications Magazine, Feb. 1991, pp. 40–42.

"Trends in Cellular and Cordless Communications", by Goodman, IEEE Communications Magazine, Jun. 1991, pp. 31–40.

"Dialing a Phone by Voice", by Pawate et al, Machine Design, Jan. 19, 1991, pp. 95–98.

"Eyes Free Dialing for Cellular Telephones", by Bendelac et al, IEEE 1991, pp. 120–125.

"Eyes Free Dialing for Cellular Telephones", by Bendelac et al, pp. 234–237.

"SPS51—A universal interface for hands–free telephony, speech recognition and speech storage in the car telephone", by Ruhl et al, Philips Telecommunication Review, vol. 48, No. 4, Dec. 1990, pp. 1–10.

"Interactive Voice Technology Applications", by Fischell et al, AT&T Technical Journal, Sep./Oct. 1990, pp. 61–76.

"A speaker independent voice dialing system for Italian in the cellular phone application", by Babini et al, CSELT Technical Reports, vol. XVIII, No. 3, Jun. 1990, pp. 201–205.

"Cellular Packet Communications", by Goodman, IEEE Transactions on Communications, vol. 38, No. 8, Aug. 1990, pp. 1272–1280.

"Isolated Word Recognition in the Mobile–Radio System: Experiments and Results", by Fissore, Signal Processing V: Theories and Applications, 1990, pp. 1207–1210.

"Packet Switching in Digital Cellular Systems", by Felix, Motorola Inc, pp. 414–418.

"Putting Speech Recognition to Work in the Telephone Network", by Lennig, Computer, Aug. 1990, pp. 35–41.

"Development of the Speaker–Dependent Voice Activated Dialing Equipment", by Sato, pp. 548–554.

"A New Speaker–Independent Voice Recognition Scheme for Voice Dialling", by Takizawa et al, IEEE 1987 p. 547–551.

"Dialing by Voice", IEEE Spectrum, Aug. 1986, p. 22.

"Voice Control of Mobile Telephones", by Helms, Speech Tech 1986, pp. 126–130.

"The 5ESS Switching System", by Martersteck et al, AT&T Technical Journal, vol. 64, No. 6, Jul., Aug. 1985, pp. 1305–1313.

"The 5ESS Switching Systems: Applications Planning", by W.R. Byre et al, AT&T Technical Journal, Jul.–Aug. 1985, pp. 1315–1336.

"The 5ESS Switching System: Architectural Overview", by D.L. Carney et al, AT&T Technical Journal, vol. 64, No. 6, Jul.–Aug. 1985, pp. 1339–1355.

"Voice Recognition in Cellular Mobile Telephones", by Thomas B. Schalk, Speech Technology, Sep./Oct. 1986, pp. 24–28.

Voice Dial Operating Guide, Voice Command System for Cellular Telephones, By–Word Technologies, Inc. 1989, pp. 1–18.

"AT&T's Conversant I Voice System", by John P. Moosemiller, Speech Technologoy, Mar./Apr. 1986, pp. 88–93.

"A PABX That Listens and Talks", Speech Technology, Jan./Feb. 1984, pp. 75–79.

"Speech Recognition Services" Projects to consder for voice input work at USWEST, from 1988 Presentation to USWEST.

Hands–free telephony, speech recognition and speech coding techniques implemented in the SPS51, by Armbruster et al, Philips Telecommunication Review, vol. 49, No. 1, Mar. 1991, pp. 19–27.

"Learned Codebook Excited Linear Predictive (LCELP) Speech Codec for Digital Cellular System", by Unno et al, NEC Res. & Develop., vol. 32, No. 4, Oct. 1991, pp. 549–556.

Abstract of Acceptance of voice system hinges on "true" recognition: Voice Ware offers speech recognition board, by Waurzyniak, Computer Software News vol., v4, Issue No. 10, Mar. 10, 1986.

"Is thought dialing next?", Telephone Engineer & Management, Apr. 1, 1991, p. 58, ISSN: 0040–263X.

* cited by examiner

FIG. 1

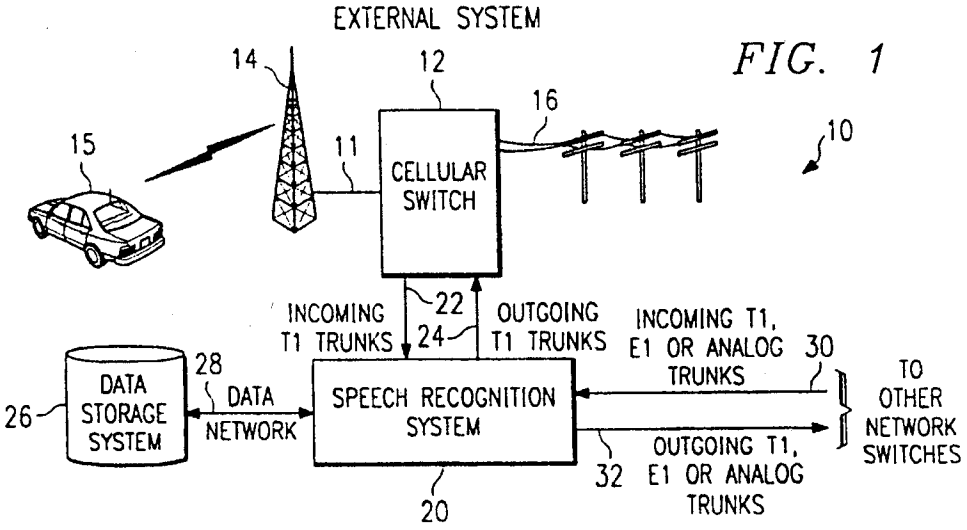
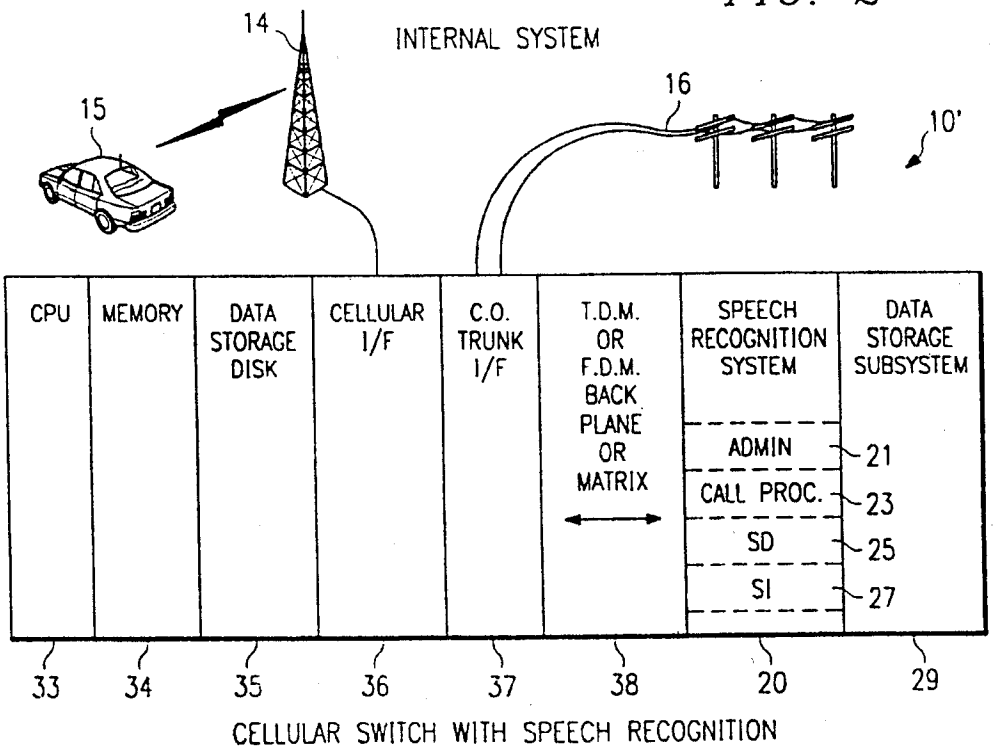


FIG. 2



CELLULAR SWITCH WITH SPEECH RECOGNITION

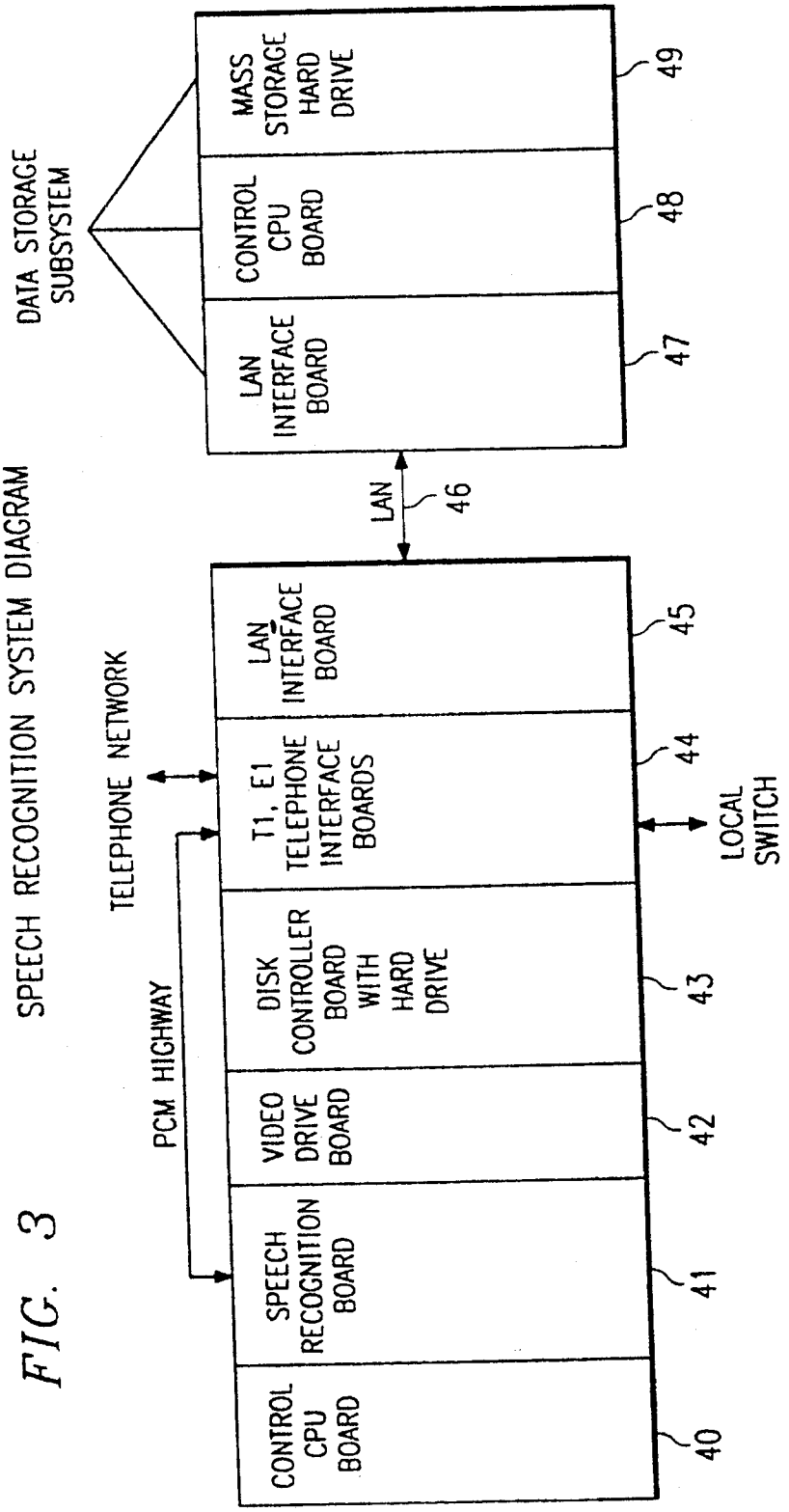

SPEECH RECOGNITION SYSTEM DIAGRAM
DATA STORAGE SUBSYSTEM
MASS STORAGE HARD DRIVE
CONTROL CPU BOARD
LAN INTERFACE BOARD
49
48
47
LAN
46
LAN INTERFACE BOARD
T1, E1 TELEPHONE INTERFACE BOARDS
DISK CONTROLLER BOARD WITH HARD DRIVE
VIDEO DRIVE BOARD
SPEECH RECOGNITION BOARD
CONTROL CPU BOARD
45
44
43
42
41
40
TELEPHONE NETWORK
PCM HIGHWAY
LOCAL SWITCH

FIG. 3

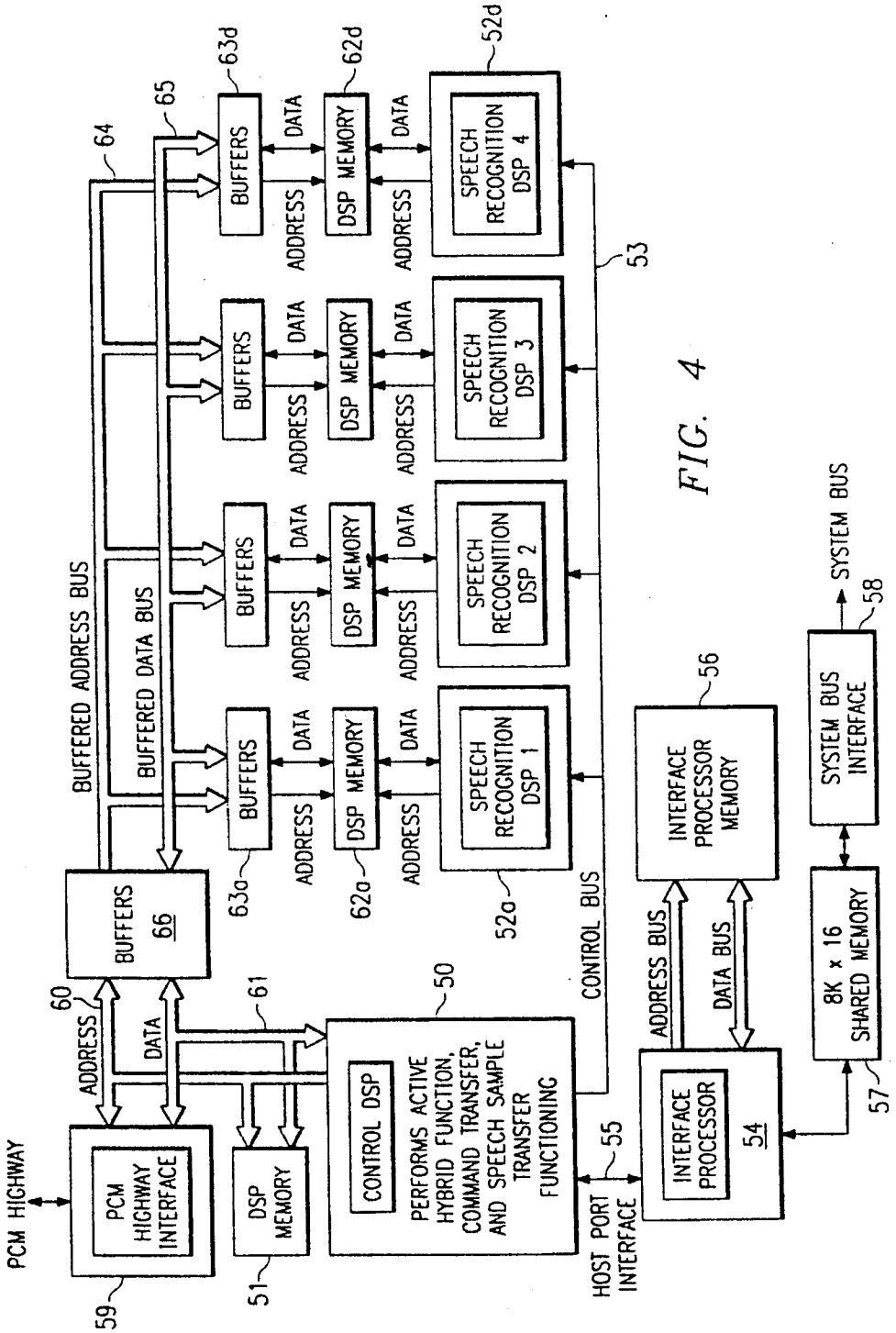
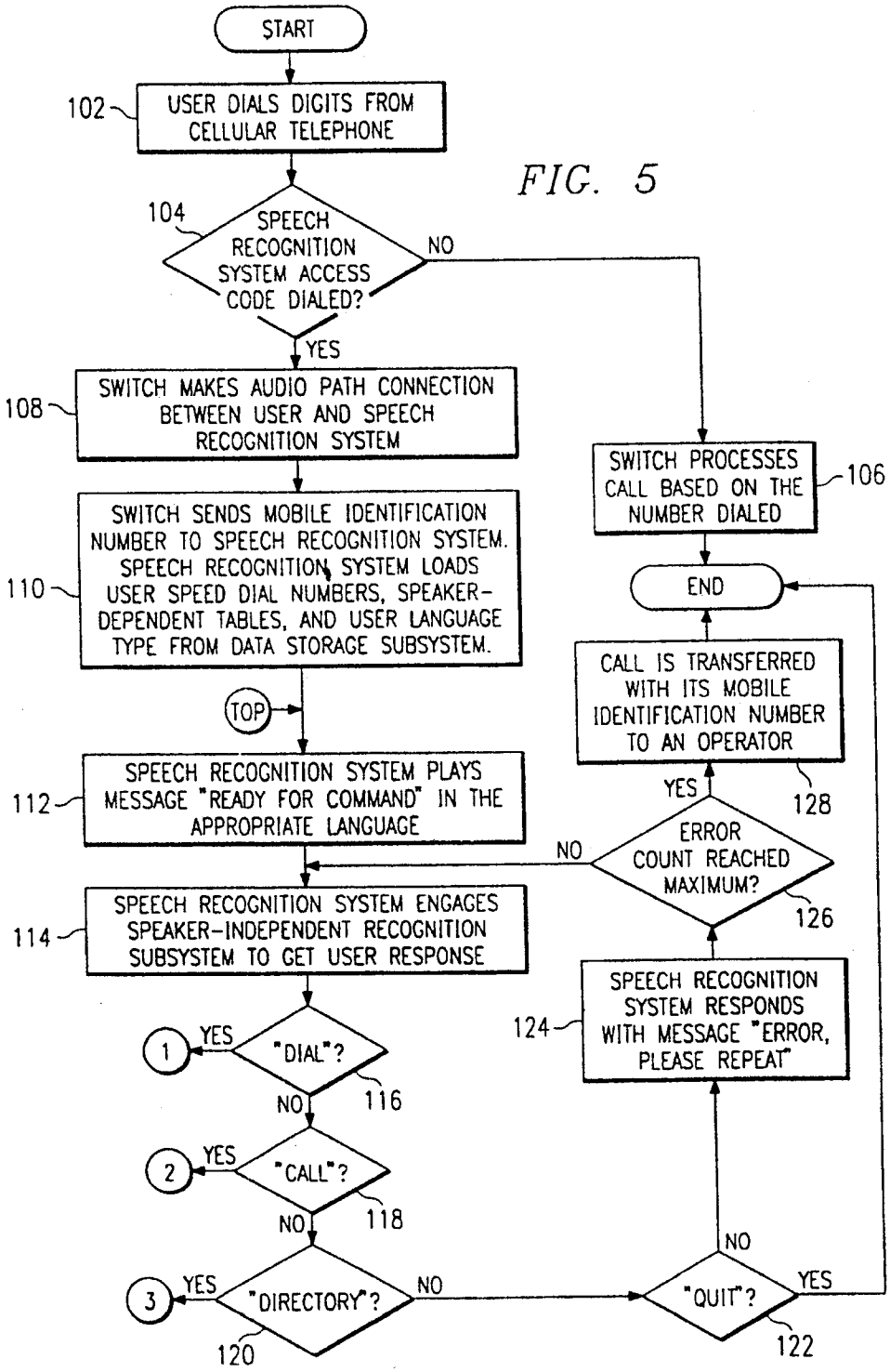

PCM HIGHWAY
PCM HIGHWAY INTERFACE
59
DSP MEMORY
51
CONTROL DSP
PERFORMS ACTIVE HYBRID FUNCTION, COMMAND TRANSFER, AND SPEECH SAMPLE TRANSFER FUNCTIONING
50
HOST PORT INTERFACE
55
INTERFACE PROCESSOR
54
ADDRESS BUS
DATA BUS
INTERFACE PROCESSOR MEMORY
56
8K x 16 SHARED MEMORY
57
SYSTEM BUS INTERFACE
58
SYSTEM BUS
CONTROL BUS
53
ADDRESS
DATA
60
61
BUFFERS 66
BUFFERED ADDRESS BUS
BUFFERED DATA BUS
64
65
63a
62a
52a
SPEECH RECOGNITION DSP 1
63d
62d
52d
SPEECH RECOGNITION DSP 4
SPEECH RECOGNITION DSP 3
SPEECH RECOGNITION DSP 2
BUFFERS
DSP MEMORY

FIG. 4

FIG. 5


START
102
USER DIALS DIGITS FROM CELLULAR TELEPHONE
104
SPEECH RECOGNITION SYSTEM ACCESS CODE DIALED?
NO
YES
108
SWITCH MAKES AUDIO PATH CONNECTION BETWEEN USER AND SPEECH RECOGNITION SYSTEM
SWITCH PROCESSES CALL BASED ON THE NUMBER DIALED
106
110
SWITCH SENDS MOBILE IDENTIFICATION NUMBER TO SPEECH RECOGNITION SYSTEM. SPEECH RECOGNITION SYSTEM LOADS USER SPEED DIAL NUMBERS, SPEAKER-DEPENDENT TABLES, AND USER LANGUAGE TYPE FROM DATA STORAGE SUBSYSTEM.
END
CALL IS TRANSFERRED WITH ITS MOBILE IDENTIFICATION NUMBER TO AN OPERATOR
TOP
112
SPEECH RECOGNITION SYSTEM PLAYS MESSAGE "READY FOR COMMAND" IN THE APPROPRIATE LANGUAGE
YES
128
NO
ERROR COUNT REACHED MAXIMUM?
126
114
SPEECH RECOGNITION SYSTEM ENGAGES SPEAKER-INDEPENDENT RECOGNITION SUBSYSTEM TO GET USER RESPONSE
124
SPEECH RECOGNITION SYSTEM RESPONDS WITH MESSAGE "ERROR, PLEASE REPEAT"
1
YES
"DIAL"?
116
NO
2
YES
"CALL"?
118
NO
NO
3
YES
"DIRECTORY"?
120
NO
"QUIT"?
YES
122

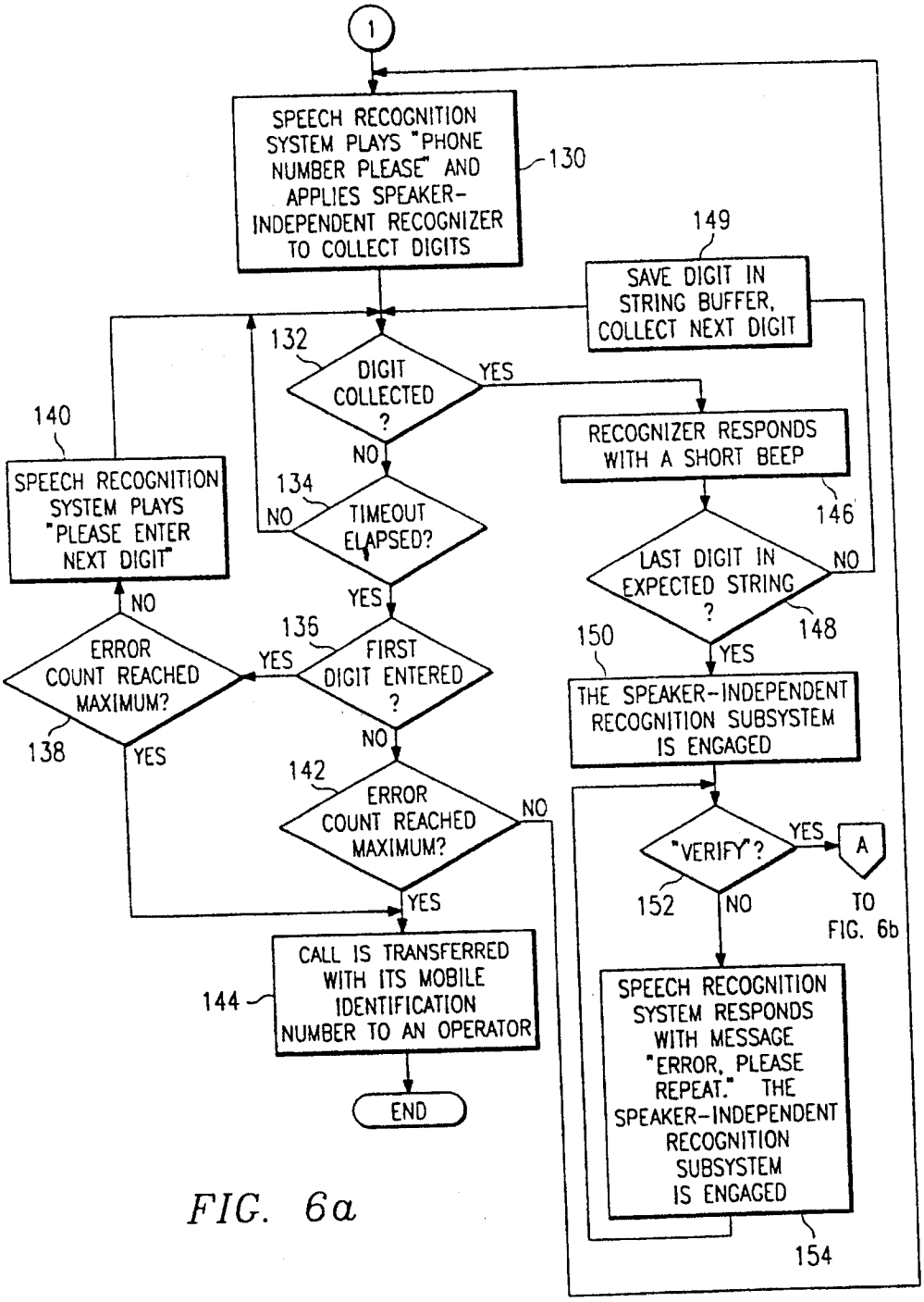

1
SPEECH RECOGNITION SYSTEM PLAYS "PHONE NUMBER PLEASE" AND APPLIES SPEAKER-INDEPENDENT RECOGNIZER TO COLLECT DIGITS
130
149
SAVE DIGIT IN STRING BUFFER, COLLECT NEXT DIGIT
132
DIGIT COLLECTED ?
YES
NO
RECOGNIZER RESPONDS WITH A SHORT BEEP
146
140
SPEECH RECOGNITION SYSTEM PLAYS "PLEASE ENTER NEXT DIGIT"
134
NO
TIMEOUT ELAPSED?
YES
LAST DIGIT IN EXPECTED STRING ?
NO
148
NO
136
ERROR COUNT REACHED MAXIMUM?
YES
FIRST DIGIT ENTERED ?
150
YES
THE SPEAKER-INDEPENDENT RECOGNITION SUBSYSTEM IS ENGAGED
138
YES
NO
142
ERROR COUNT REACHED MAXIMUM?
NO
"VERIFY"?
YES
A
152
NO
TO FIG. 6b
YES
CALL IS TRANSFERRED WITH ITS MOBILE IDENTIFICATION NUMBER TO AN OPERATOR
144
SPEECH RECOGNITION SYSTEM RESPONDS WITH MESSAGE "ERROR, PLEASE REPEAT." THE SPEAKER-INDEPENDENT RECOGNITION SUBSYSTEM IS ENGAGED
154
END

FIG. 6a

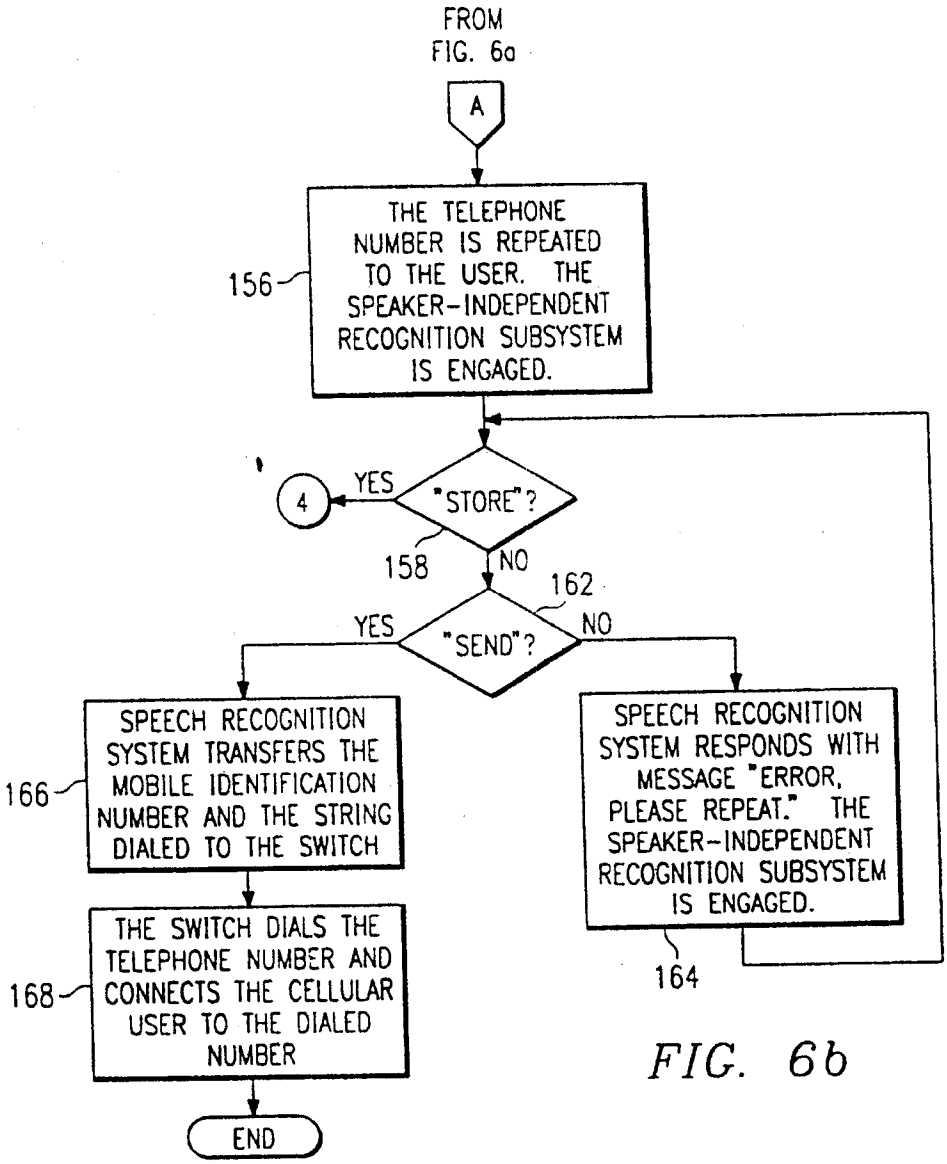

FROM FIG. 6a
A
THE TELEPHONE NUMBER IS REPEATED TO THE USER. THE SPEAKER-INDEPENDENT RECOGNITION SUBSYSTEM IS ENGAGED.
156
"STORE"?
YES
4
158
NO
162
"SEND"?
YES
NO
SPEECH RECOGNITION SYSTEM TRANSFERS THE MOBILE IDENTIFICATION NUMBER AND THE STRING DIALED TO THE SWITCH
166
THE SWITCH DIALS THE TELEPHONE NUMBER AND CONNECTS THE CELLULAR USER TO THE DIALED NUMBER
168
END
SPEECH RECOGNITION SYSTEM RESPONDS WITH MESSAGE "ERROR, PLEASE REPEAT." THE SPEAKER-INDEPENDENT RECOGNITION SUBSYSTEM IS ENGAGED.
164

FIG. 6b

FIG. 7a

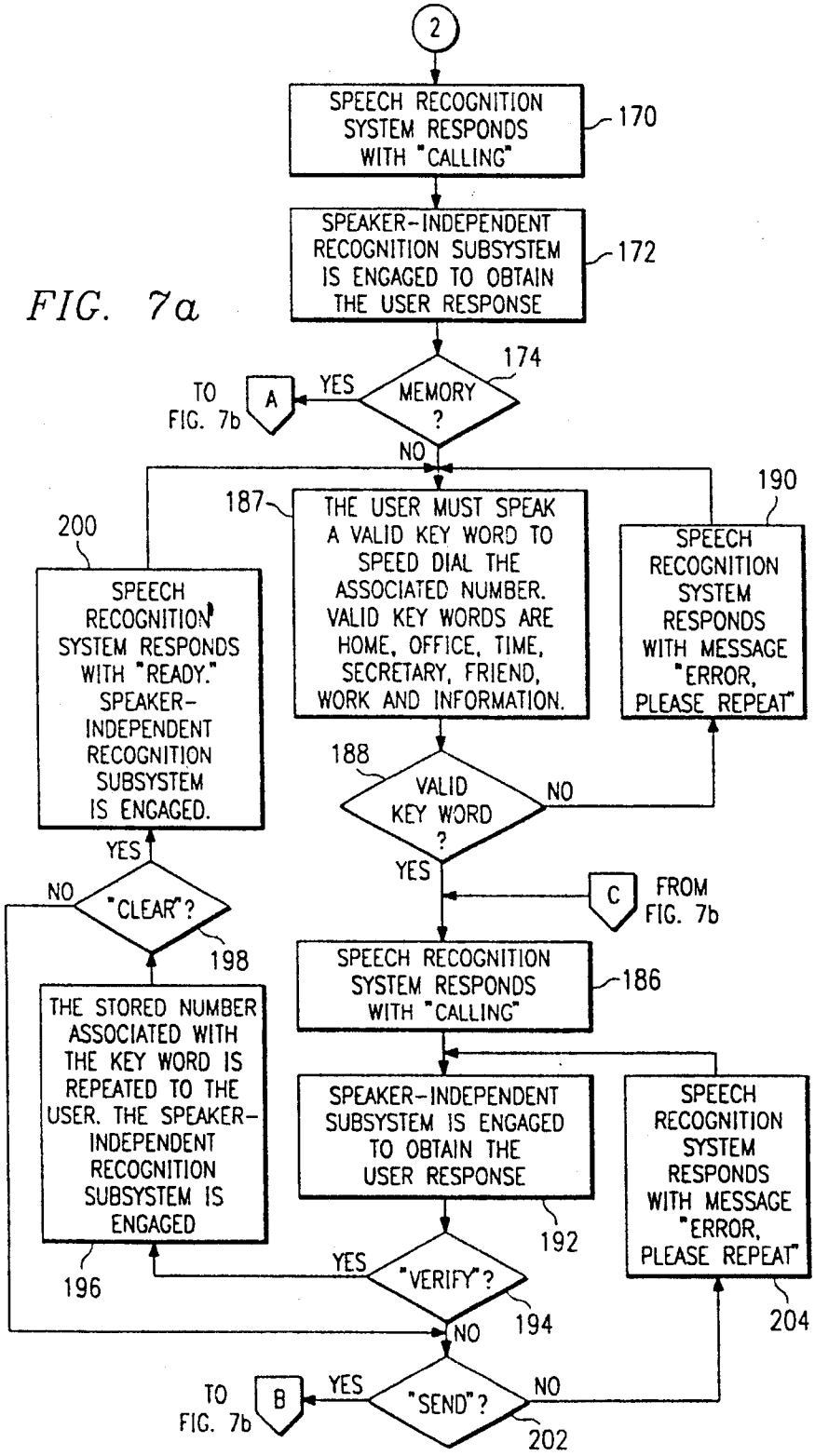
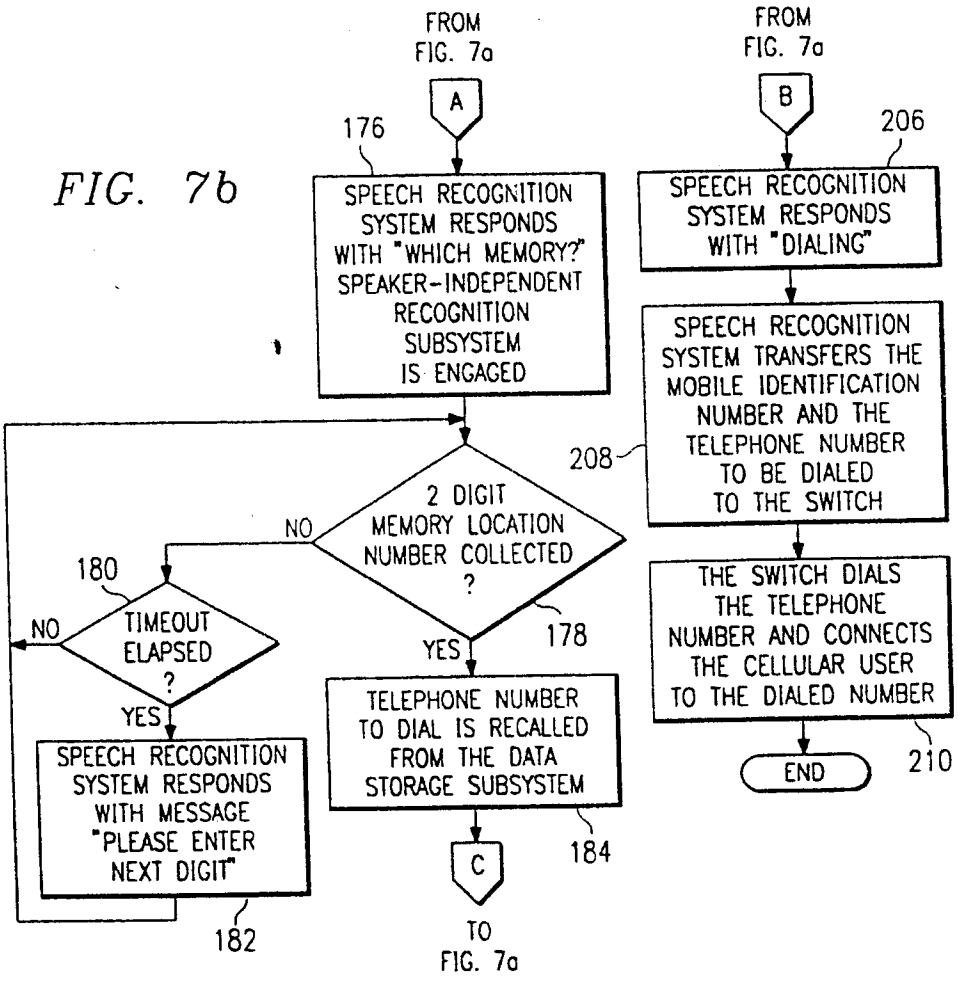

2
SPEECH RECOGNITION SYSTEM RESPONDS WITH "CALLING"
170
SPEAKER-INDEPENDENT RECOGNITION SUBSYSTEM IS ENGAGED TO OBTAIN THE USER RESPONSE
172
MEMORY ?
174
YES
TO FIG. 7b
A
NO
187
THE USER MUST SPEAK A VALID KEY WORD TO SPEED DIAL THE ASSOCIATED NUMBER. VALID KEY WORDS ARE HOME, OFFICE, TIME, SECRETARY, FRIEND, WORK AND INFORMATION.
190
SPEECH RECOGNITION SYSTEM RESPONDS WITH MESSAGE "ERROR, PLEASE REPEAT"
200
SPEECH RECOGNITION SYSTEM RESPONDS WITH "READY." SPEAKER-INDEPENDENT RECOGNITION SUBSYSTEM IS ENGAGED.
188
VALID KEY WORD ?
NO
YES
YES
NO
"CLEAR"?
198
C
FROM FIG. 7b
SPEECH RECOGNITION SYSTEM RESPONDS WITH "CALLING"
186
THE STORED NUMBER ASSOCIATED WITH THE KEY WORD IS REPEATED TO THE USER. THE SPEAKER-INDEPENDENT RECOGNITION SUBSYSTEM IS ENGAGED
SPEAKER-INDEPENDENT SUBSYSTEM IS ENGAGED TO OBTAIN THE USER RESPONSE
SPEECH RECOGNITION SYSTEM RESPONDS WITH MESSAGE "ERROR, PLEASE REPEAT"
192
204
196
YES
"VERIFY"?
194
NO
TO FIG. 7b
B
YES
"SEND"?
NO
202

FIG. 7b


FROM FIG. 7a
A
176
SPEECH RECOGNITION SYSTEM RESPONDS WITH "WHICH MEMORY?" SPEAKER-INDEPENDENT RECOGNITION SUBSYSTEM IS ENGAGED
2 DIGIT MEMORY LOCATION NUMBER COLLECTED ?
NO
YES
178
180
TIMEOUT ELAPSED ?
NO
YES
SPEECH RECOGNITION SYSTEM RESPONDS WITH MESSAGE "PLEASE ENTER NEXT DIGIT"
182
TELEPHONE NUMBER TO DIAL IS RECALLED FROM THE DATA STORAGE SUBSYSTEM
184
C
TO FIG. 7a
FROM FIG. 7a
B
206
SPEECH RECOGNITION SYSTEM RESPONDS WITH "DIALING"
208
SPEECH RECOGNITION SYSTEM TRANSFERS THE MOBILE IDENTIFICATION NUMBER AND THE TELEPHONE NUMBER TO BE DIALED TO THE SWITCH
THE SWITCH DIALS THE TELEPHONE NUMBER AND CONNECTS THE CELLULAR USER TO THE DIALED NUMBER
END
210

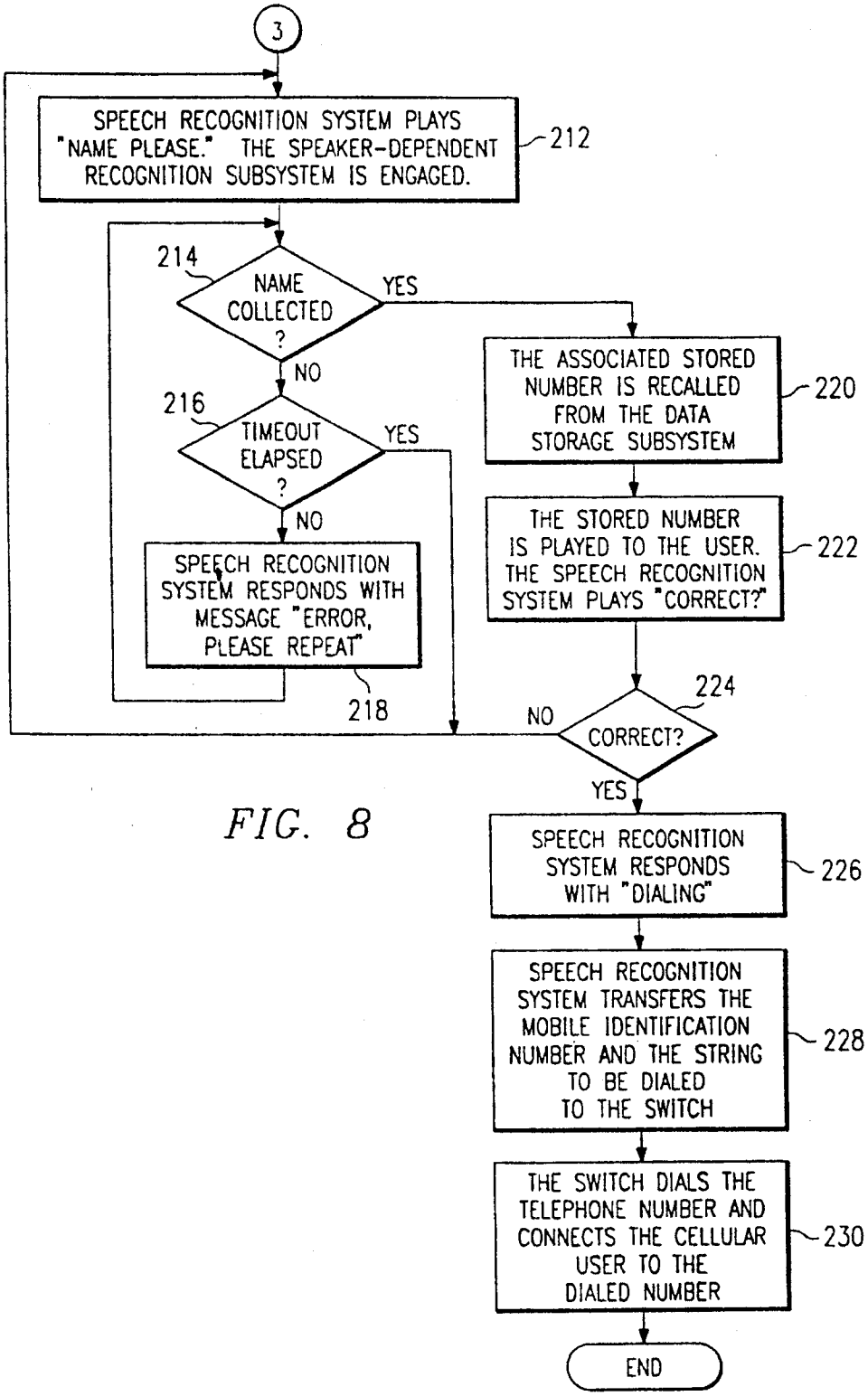

3
SPEECH RECOGNITION SYSTEM PLAYS "NAME PLEASE." THE SPEAKER-DEPENDENT RECOGNITION SUBSYSTEM IS ENGAGED.
212
214
NAME COLLECTED ?
YES
NO
216
TIMEOUT ELAPSED ?
YES
NO
SPEECH RECOGNITION SYSTEM RESPONDS WITH MESSAGE "ERROR, PLEASE REPEAT"
218
THE ASSOCIATED STORED NUMBER IS RECALLED FROM THE DATA STORAGE SUBSYSTEM
220
THE STORED NUMBER IS PLAYED TO THE USER. THE SPEECH RECOGNITION SYSTEM PLAYS "CORRECT?"
222
224
NO
CORRECT?
YES
SPEECH RECOGNITION SYSTEM RESPONDS WITH "DIALING"
226
SPEECH RECOGNITION SYSTEM TRANSFERS THE MOBILE IDENTIFICATION NUMBER AND THE STRING TO BE DIALED TO THE SWITCH
228
THE SWITCH DIALS THE TELEPHONE NUMBER AND CONNECTS THE CELLULAR USER TO THE DIALED NUMBER
230
END

FIG. 8

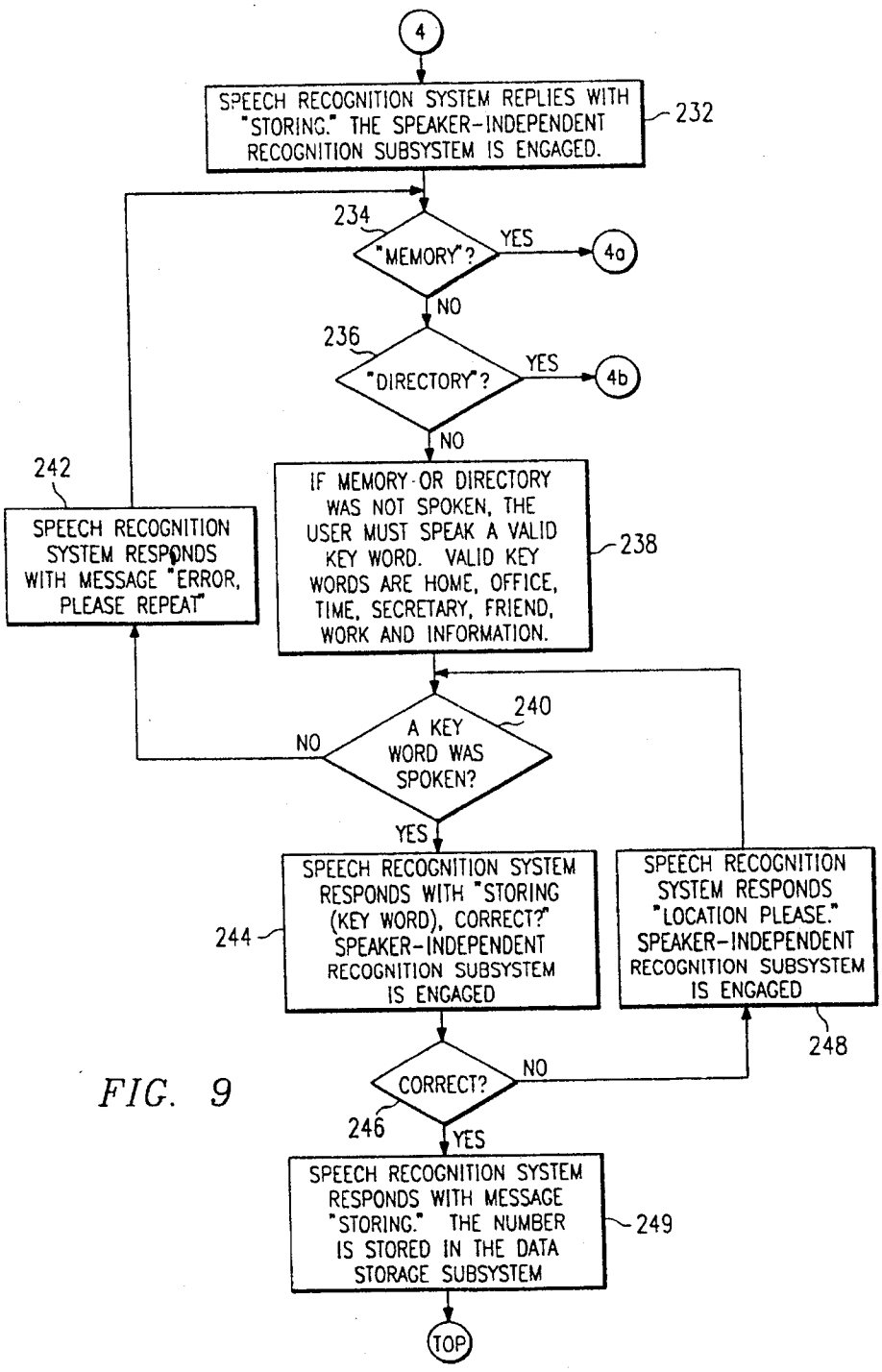

4
SPEECH RECOGNITION SYSTEM REPLIES WITH "STORING." THE SPEAKER-INDEPENDENT RECOGNITION SUBSYSTEM IS ENGAGED.
232
234
"MEMORY"?
YES
4a
NO
236
"DIRECTORY"?
YES
4b
NO
242
SPEECH RECOGNITION SYSTEM RESPONDS WITH MESSAGE "ERROR, PLEASE REPEAT"
IF MEMORY OR DIRECTORY WAS NOT SPOKEN, THE USER MUST SPEAK A VALID KEY WORD. VALID KEY WORDS ARE HOME, OFFICE, TIME, SECRETARY, FRIEND, WORK AND INFORMATION.
238
240
A KEY WORD WAS SPOKEN?
NO
YES
SPEECH RECOGNITION SYSTEM RESPONDS WITH "STORING (KEY WORD), CORRECT?" SPEAKER-INDEPENDENT RECOGNITION SUBSYSTEM IS ENGAGED
244
SPEECH RECOGNITION SYSTEM RESPONDS "LOCATION PLEASE." SPEAKER-INDEPENDENT RECOGNITION SUBSYSTEM IS ENGAGED
248
CORRECT?
NO
246
YES
SPEECH RECOGNITION SYSTEM RESPONDS WITH MESSAGE "STORING." THE NUMBER IS STORED IN THE DATA STORAGE SUBSYSTEM
249
TOP

FIG. 9

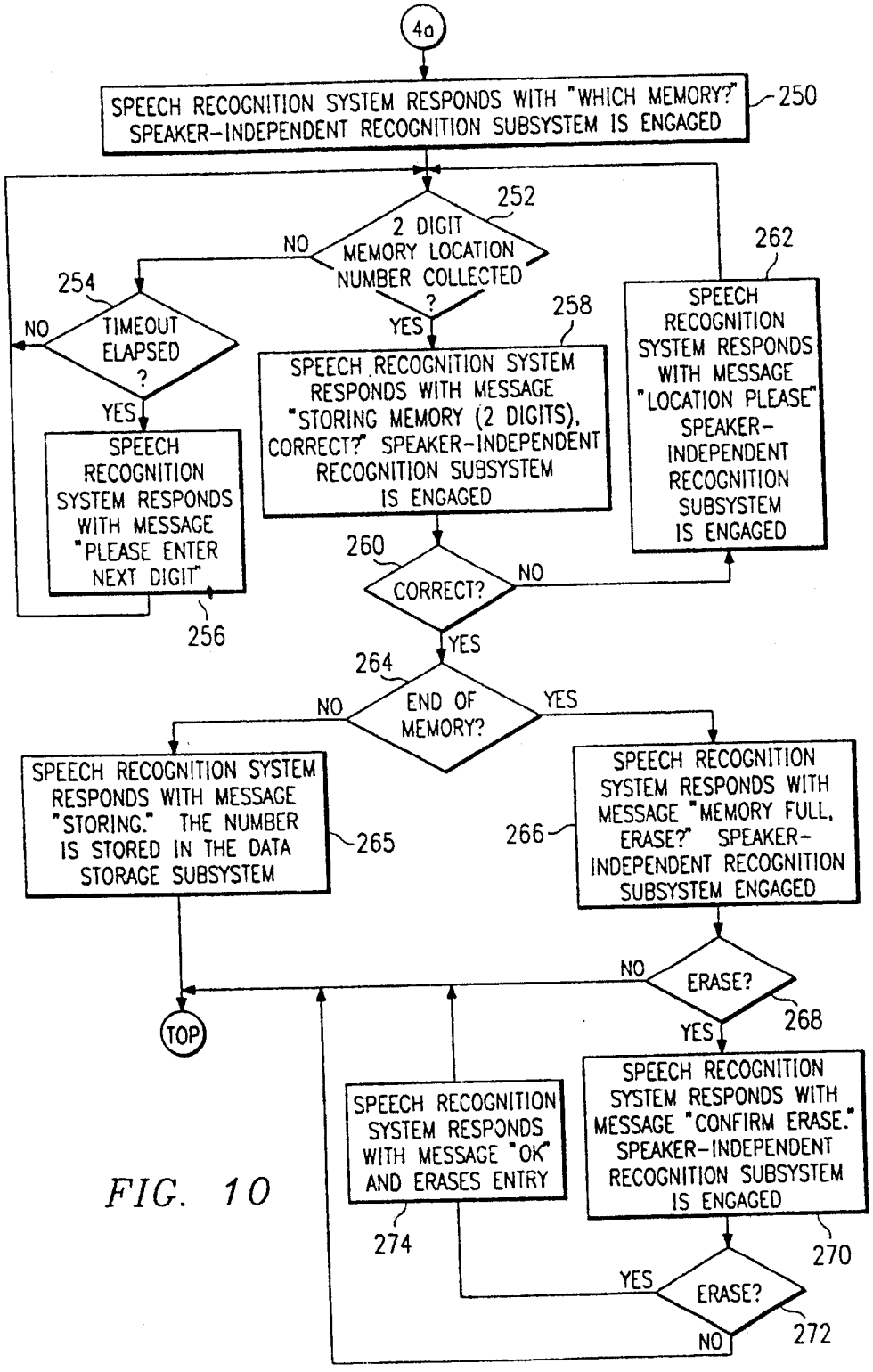

4a
SPEECH RECOGNITION SYSTEM RESPONDS WITH "WHICH MEMORY?" SPEAKER-INDEPENDENT RECOGNITION SUBSYSTEM IS ENGAGED
250
252
2 DIGIT MEMORY LOCATION NUMBER COLLECTED ?
NO
YES
254
TIMEOUT ELAPSED ?
NO
YES
SPEECH RECOGNITION SYSTEM RESPONDS WITH MESSAGE "PLEASE ENTER NEXT DIGIT"
256
258
SPEECH RECOGNITION SYSTEM RESPONDS WITH MESSAGE "STORING MEMORY (2 DIGITS), CORRECT?" SPEAKER-INDEPENDENT RECOGNITION SUBSYSTEM IS ENGAGED
260
CORRECT?
NO
YES
262
SPEECH RECOGNITION SYSTEM RESPONDS WITH MESSAGE "LOCATION PLEASE" SPEAKER-INDEPENDENT RECOGNITION SUBSYSTEM IS ENGAGED
264
END OF MEMORY?
NO
YES
SPEECH RECOGNITION SYSTEM RESPONDS WITH MESSAGE "STORING." THE NUMBER IS STORED IN THE DATA STORAGE SUBSYSTEM
265
266
SPEECH RECOGNITION SYSTEM RESPONDS WITH MESSAGE "MEMORY FULL, ERASE?" SPEAKER-INDEPENDENT RECOGNITION SUBSYSTEM ENGAGED
NO
ERASE?
YES
268
TOP
SPEECH RECOGNITION SYSTEM RESPONDS WITH MESSAGE "OK" AND ERASES ENTRY
274
SPEECH RECOGNITION SYSTEM RESPONDS WITH MESSAGE "CONFIRM ERASE." SPEAKER-INDEPENDENT RECOGNITION SUBSYSTEM IS ENGAGED
270
YES
ERASE?
NO
272

FIG. 10

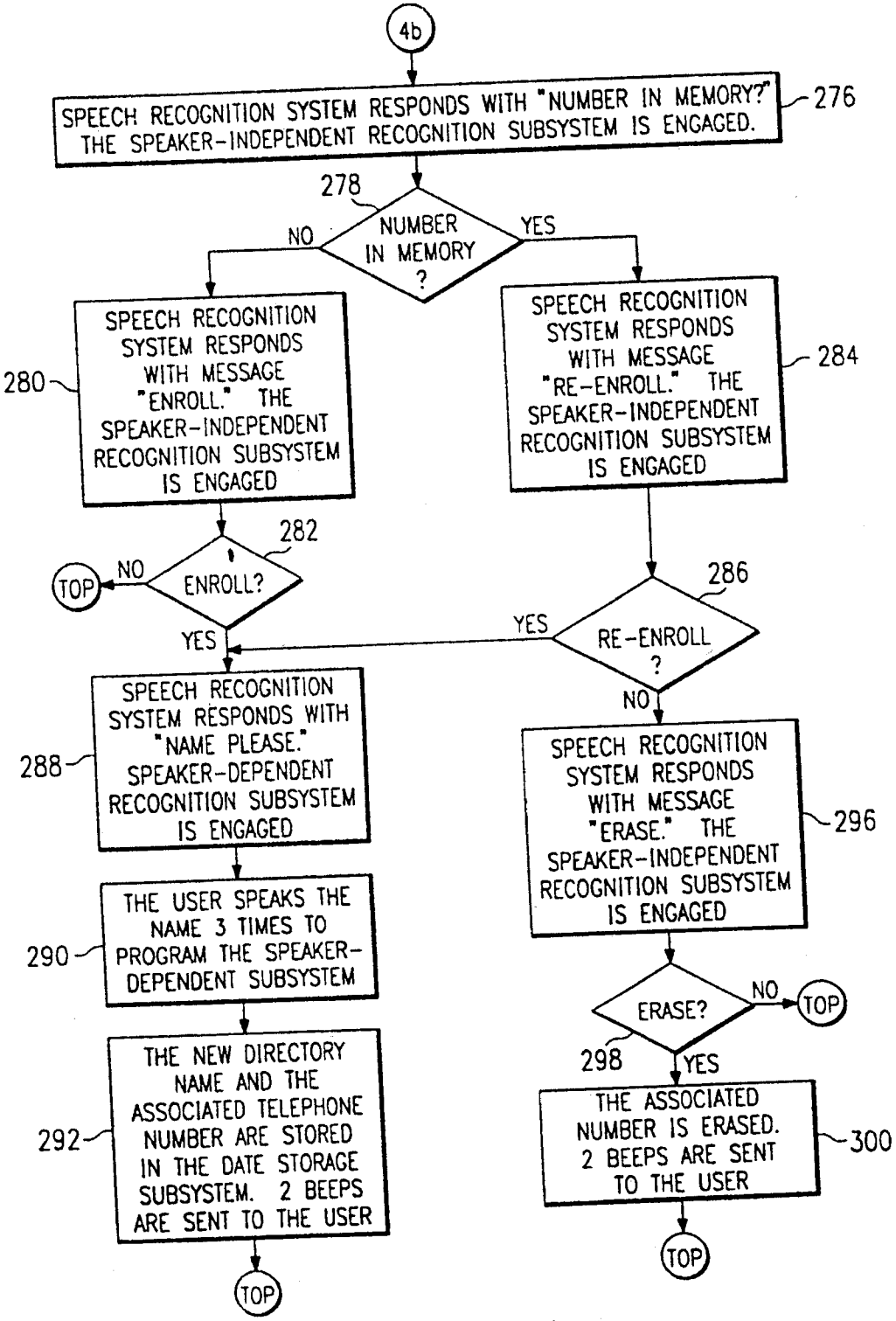

4b
SPEECH RECOGNITION SYSTEM RESPONDS WITH "NUMBER IN MEMORY?" THE SPEAKER-INDEPENDENT RECOGNITION SUBSYSTEM IS ENGAGED.
276
278
NUMBER IN MEMORY ?
NO
YES
SPEECH RECOGNITION SYSTEM RESPONDS WITH MESSAGE "ENROLL." THE SPEAKER-INDEPENDENT RECOGNITION SUBSYSTEM IS ENGAGED
280
SPEECH RECOGNITION SYSTEM RESPONDS WITH MESSAGE "RE-ENROLL." THE SPEAKER-INDEPENDENT RECOGNITION SUBSYSTEM IS ENGAGED
284
282
ENROLL?
NO
TOP
YES
286
RE-ENROLL ?
YES
NO
SPEECH RECOGNITION SYSTEM RESPONDS WITH "NAME PLEASE." SPEAKER-DEPENDENT RECOGNITION SUBSYSTEM IS ENGAGED
288
SPEECH RECOGNITION SYSTEM RESPONDS WITH MESSAGE "ERASE." THE SPEAKER-INDEPENDENT RECOGNITION SUBSYSTEM IS ENGAGED
296
THE USER SPEAKS THE NAME 3 TIMES TO PROGRAM THE SPEAKER-DEPENDENT SUBSYSTEM
290
ERASE?
NO
TOP
298
YES
THE NEW DIRECTORY NAME AND THE ASSOCIATED TELEPHONE NUMBER ARE STORED IN THE DATE STORAGE SUBSYSTEM. 2 BEEPS ARE SENT TO THE USER
292
THE ASSOCIATED NUMBER IS ERASED. 2 BEEPS ARE SENT TO THE USER
300
TOP
TOP

FIG. 11

SPEECH RECOGNITION SYSTEM FOR ELECTRONIC SWITCHES IN A NON-WIRELINE COMMUNICATIONS NETWORK

CROSS REFERENCE TO RELATED APPLICATION

This application is a continuation of application Ser. No. 08/914,440, filed Aug. 19, 1997, now U.S. Pat. No. 6,157,848, which is a continuation of application Se. No. 08/216,009 filed Mar. 22, 1994 now U.S. Pat. No. 5,659,597, which is a continuation of application Ser. No. 07,867,873, filed Apr. 13, 1992, now U.S. Pat. No. 5,297,183.

TECHNICAL FIELD

This invention is related to telecommunications systems, and more particularly to an electronic digital signal processor-controlled telecommunications system for the recognition of spoken commands and for the directing of telephone calls based on spoken commands.

BACKGROUND OF THE INVENTION

It is well-known that many of the safety hazards of cellular telephone use could be alleviated by utilizing automatic speech recognition. While telephone-based speech recognition systems are known, cellular voice dialing over a mobile telephone exchange ("MTX") presents significant challenges for two basic reasons. First, the recognition technology must accommodate a tremendous range of both remotely-mounted and hand-held microphone types. Second, the signal may be band-limited and degraded in transmission to the MTX where the recognition system will be located. Voice-controlled dialers of the prior art, such as taught in U.S. Pat. No. 4,853,953 to Fujisaki, have not been successfully implemented in the cellular environment.

There is therefore a need for voice recognition systems for use in the cellular, satelite and and personal communications network environments that overcome these and other problems of the prior art and that facilitate the use of voice-dialing and other safety and convenience features.

BRIEF SUMMARY OF THE INVENTION

It is therefore an object of the present invention to describe an implementation of a speech recognition system in a cellular or personal communications network environment.

It is a further object of the invention to describe a speech recognition system for use at a mobile telephone exchange (MTX) of a cellular or personal communications network. The placement of the speech recognition system at the MTX significantly reduces cost and increases reliability by enabling the switch vendor to install and maintain the system in conjunction with the cellular switch.

It is another object of the invention to describe a cellular voice dialing system for use in or in conjunction with an MTX of a cellular network.

It is still another object of the invention to use voice recognition techniques to secure access to a cellular or personal communications network.

Another object of the invention is to provide for combined use of speaker-dependent and speaker-independent voice recognition and speaker verification techniques in an MTX of a cellular or personal communications telephone network.

These and other objects of the invention are provided in an advanced system for the recognizing of spoken commands over the cellular telephone or any personal communications (i.e., any non-wireline) network. In the cellular application, for example, a Speech Recognition System interconnects either internally with or as an external peripheral to a cellular telecommunications MTX switch. The Speech Recognition System includes an administrative subsystem, a call processing subsystem, a speaker-dependent recognition subsystem, a speaker-independent recognition subsystem, and a data storage subsystem. The Speech Recognition System also allows for increased efficiency in the cellular telephone network by integrating with the switch or switches as a shared resource. The administrative subsystem of the Speech Recognition System is used to keep statistical logs of pertinent call information. Prerecorded instructional messages are stored in the memory of the call processing subsystem for instructing a user on his or her progress in using the system. The speaker-independent recognition subsystem allows the user to interact with the system employing non-user specific functions. User specific functions are controlled with the speaker-dependent recognition subsystem. User specific attributes collected by the recognition subsystems are stored in the data storage subsystem.

The foregoing has outlined some of the more pertinent objects of the present invention. These objects should be construed to be merely illustrative of some of the more prominent features and applications of the invention. Many other beneficial results can be attained by applying the disclosed invention in a different manner or modifying the invention as will be described. Accordingly, other objects and a fuller understanding of the invention may be had by referring to the following Detailed Description of the preferred embodiment.

BRIEF DESCRIPTION OF THE DRAWINGS

For a more complete understanding of the present invention and the advantages thereof, reference should be made to the following Detailed Description taken in connection with the accompanying drawings in which:

FIG. **1** is a block diagram of a cellular telephone network incorporating an external switch-based Speech Recognition System according to the present invention;

FIG. **2** is a block diagram of a cellular telephone network incorporating an internal switch-based Speech Recognition System;

FIG. **3** is an upper level block diagram of the Speech Recognition System and data storage subsystem of FIG. **1**;

FIG. **4** is a detailed component diagram of the speech recognizer board of the Speech Recognition System of FIGS. **2** and **3**;

FIG. **5** is a flowchart showing a Control routine used in the Speech Recognition System;

FIG. **6** is a flowchart of the Dial routine of the invention used by the subscriber to voice dial a telephone number that has been previously-stored by the subscriber;

FIG. **7** is a flowchart of the Call routine used by the subscriber to voice dial a speaker-independent telephone number;

FIG. **8** is a flowchart of the Directory Dialing routine used by the subscriber to recall a previously-stored number using a speaker-dependent directory name;

FIG. **9** is a flowchart of a Store routine used to store subscriber-specific destination number information;

FIG. **10** is a flowchart of a Memory routine used to store speed-dial numbers associated with specific memory locations; and

FIG. **11** is a flowchart of a Directory Storage routine used to store speaker-dependent directory names and associated telephone numbers for use by the subscriber to speed-dial destination numbers based on the associated directory names.

Similar reference characters refer to similar parts or steps throughout the several drawings.

DETAILED DESCRIPTION

FIG. **1** is a block diagram of a cellular telephone network incorporating an external switch-based Speech Recognition System according to the present invention. Although the following description is specifically related to use of the Speech Recognition System at or in conjunction with an MTX of a cellular network, it should be appreciated that the System also interconnects either internally with or as an external peripheral to a personal communications network. Indeed, the principles of the invention are applicable to any cellular-like network application, i.e., where a non-wireline communications network is employed for mobile, satelite, portable or personal communications. The Speech Recognition System can also be used as a shared resource through integration with a plurality of such non-wireline communications networks.

Referring now to FIG. **1**, an exemplary telephone network is a cellular network **10** having a mobile telephone exchange (MTX) switch **12** connected to a transceiver and an antenna **14**. The transceiver is located in each cell of the cellular network and communicates with the MTX to effect transmission and reception of signals to and from the mobile telephone located in vehicle **15**. The transceiver is typically connected to the MTX via a leased or dedicated network line **11**. The MTX **12** is typically connected to the land-based destinations via telephone network **16**.

A cellular mobile telecommunications system connects mobile telecommunications customers, each having a mobile unit, to land-based customers served by a telephone network. Incoming and outgoing calls are routed through a mobile telecommunications switching office connected to a group of cell sites that communicate with mobile units. The mobile telecommunications switching office includes a mobile telephone switching exchange (MTX) for routing the calls between the mobile units and the telephone network. In a typically mobile cellular communications system, there are usually many cells per MTX and several MTX's per system. As used herein, "mobile telecommunications system" refers to cellular, satelite and personal communications network environments.

Each cellular telephone is uniquely identified by two numbers: a serial number ("ESN") encoded in the phone by its manufacturer, and a mobile identification number ("MIN"), which is the cellular telephone number programmed in by the cellular service provider. The service provider operates the MTX **12** and keeps a database of all MIN's (and their associated ESN's). Each time a call is placed by the cellular telephone, the service provider verifies whether the MIN and the associated ESN are authorized. If the MIN and ESN of a cellular phone are not recognized (and the area code or NPA of the MIN indicates that the phone number is outside of the provider's service area), the provider normally allows the call to proceed at least one time. If billing authorization cannot later be verified, however, the MIN and serial number are then placed on an exception list. Subsequent attempts to use the MIN will then be rejected.

According to one embodiment of the invention as shown in FIG. **1**, a Speech Recognition System **20** is connected as an external peripheral to the MTX through a set of preferably digital trunk lines. Set **22** is used for incoming signals and set **24** is used for outgoing signals. Other types of signaling, such as CEPT E1 or analog, may also be used besides T1. The Speech Recognition System **20** is connected to a dedicated data storage subsystem **26** through a data network **28**. The data storage subsystem is used to store recognition data derived from the subscribers to the voice dialing service as will be described. The Speech Recognition System **20** may be integrated with one or more switches (whether or not cellular) for use as a shared resource via incoming and outgoing trunk sets **30** and **32**.

Referring now to FIG. **2**, a block diagram shows the cellular telephone network **10'** with the Speech Recognition System **20** interconnected internally to the MTX. This is the preferred embodiment of the invention. The hierarchical architecture of the cellular switch includes the central processing unit **33**, memory **34**, data storage disk **35**, cellular interface **36**, central office trunk interface **37** and a backplane or switching matrix **38**.

The Speech Recognition System **20** includes a number of functional subsystems: an administrative subsystem **21**, a call processing subsystem **23**, a speaker-dependent recognition subsystem **25**, a speaker-independent recognition subsystem **27**, and the data storage subsystem **29** (which corresponds to the storage system **26** of FIG. **1**). The administrative subsystem **21** of the Speech Recognition System is used to keep statistical logs of pertinent call information. Pre-recorded instructional messages are stored in the memory of the call processing subsystem **23** for instructing a user on his or her progress in using the system. The speaker-independent recognition subsystem **27** allows the user to interact with the system employing non-user specific functions. User specific functions are controlled with the speaker-dependent recognition subsystem **25**. User specific attributes collected by the recognition subsystems are stored in the data storage subsystem **29**.

FIG. **3** is an upper level block diagram of the Speech Recognition System of FIG. **1**. The hierarchical architecture of the System **20** comprises a control central processing unit **40**, a speech recognizer board **41**, a video drive circuit board **42**, a disk drive controller board **43** with associated hard disk drive, telephone interface circuit boards **44**, and a local area network ("LAN") interface board **45**. A local area network **46** connects these components of the System to the data storage subsystem **29**, which comprises a LAN interface board **47**, a CPU control board **48** and a mass storage hard drive device **49**. As also seen in FIG. **3**, a pulse code modulation ("PCM") highway connects the telephone interface boards **44** to the speech recognition board **41**. The MTX is also connected to the System **20** through the interface boards **44**.

The CPU **40** and associated control programs function as the system administrative subsystem **21** of FIG. **2**. The recognizer board **41** acts as the recognition call processing subsystem **23** and the recognition subsystems **25** and **27**. Command and control functions are communicated to the recognizer board **41** from the system administrative processor **40** to the recognition call processing subsystem through a system bus. Responses received by the administrative processor from the recognition subsystem indicate primary and secondary recognition responses, error codes and command acknowledgements.

Referring now to FIG. **4**, a more detailed schematic diagram is shown of the recognizer board **41** of FIG. **3** and modules **25** and **27** of FIG. **2**. The recognizer includes a

control digital signal processor ("DSP") **50** having an associated memory **51** for supporting control programs and data. The control DSP **50** in this embodiment of the invention controls a plurality of speech recognition digital signal processors **52***a* . . . **52***d* via a control bus **53**. The control DSP **50** also connects to an interface processor **54** via a host port interface **55**. Interface processor **54** has its own processor memory **56**. A buffer **57** interconnects the interface processor to a system bus interface circuit **58** that interconnects the recognizer board to the system bus.

The PCM highway of FIG. **3** or backplane of FIG. **2** is connected to a PCM highway interface circuit **59** that supplies speech sample information to the control DSP **50** and memory **51** via address and data bases **60** and **61**. Each of the speech recognition DSP's **52***a* . . . **52***d* has an associated memory **62***a* . . . **62***d* and storage buffer **63***a* . . . **63***d*. Buffered address and data buses **64** and **65** interconnect to the address and data buses **60** and **61** through the buffers **66**.

The interface processor **64** converts batch system commands or command blocks received through the system bus interface **58** into singular commands for the control DSP **50**. Similarly, singular responses from the control DSP **50** are buffered by the interface processor **54** and are sent to the administrative processor (of FIGS. **2** and **3**) in code blocks to increase overall system efficiency.

All telephone channel signaling and PCM highway sample transfer is handled by the control DSP **50** through the PCM highway interface **59**. The control DSP **50** is also used to send samples and commands to the four speech recognition DSP's **52***a* . . . **52***d*. Additionally, the DSP's **52** can function to stop outgoing messages based on detection of certain incoming speech energy.

The system data flow is such that recognition commands and responses are sent to and from the administrative processor and the speech recognition DSP's **52** through the control DSP **50**, the interface processor **54**, and the system bus interface **58**. Speech samples are received by the speech recognition DSP's **52** through the PCM highway interface **59** and the control DSP **50**.

The present invention facilitates the implementation of voice-dialing in a cellular telephone or other personal communications network environment. When a user "subscribes" to the service (e.g., with the MTX service provider), it is desirable that certain speech data be collected from the subscriber for security purposes. Thus, upon subscription, the user normally will be asked to provide his or her native language (e.g., English, Spanish, French, etc.), a personal identification number, and personal information related specifically to the subscriber such as a social security number or date of birth. This latter information may be used in an automated query process as will be described to prevent fraudulent use of the cellular or personal telephone network. Once the user information is activated in the system, the user may place or receive telephone calls. Initially, the user will be required to enter speed-dial numbers and their associated directory names.

Referring now to FIG. **5**, a functional flowchart is provided describing the basic control functions of the Speech Recognition System according to the present-invention. The routine begins at step **102** when the user dials digits from the cellular telephone. At step **104**, a test is performed to determine if a Speech Recognition System access code has been dialed. If not, the cellular switch processes the call based on the number dialed at step **106** and the routine ends. If the result of the test at step **104** is positive, the routine

continues at step **108** during which the switch makes an audio path connection between the user and the Speech Recognition System. At step **110**, the switch sends the user's mobile identification number ("MIN") to the Speech Recognition System. As noted above, the MIN is a unique number associated with a given cellular telephone that is available to the switch each time a telephone call is placed.

According to the invention, each user who subscribes to the service will have prerecorded a list of destination numbers. At step **110**, these speed-dial numbers, along with speaker-dependent templates and user language type data, are retrieved from the data storage subsystem. As noted above, the data storage subsystem stores such data at predetermined locations that are preferably accessed by the MIN. The routine then continues at step **112** with the Speech Recognition System prompting the user that it is "Ready For Command" or the like. The command is made in the language as determined by the user language type data retrieved at step **110**. At step **114**, the Speech Recognition Systems engages the speaker-independent recognition subsystem to obtain the user response. Depending on the response, one of several different subroutines follow.

If the user states and the system recognizes a "Dial" command, control is passed to the Dial Routine of FIG. **6**. In particular, a test is made at step **116** to determine if the Dial command is recognized. If so, control is transferred to the routine of FIG. **6**. If the response to the test at step **116** is negative, a test is made at step **118** to determine if a "Call" command has been spoken and recognized. If the system recognizes a "Call" Command, control is passed to the Call Routine of FIG. **7**. If the response to the test at step **118** is negative, a test is made at step **120** to determine if a "Directory" command has been spoken and recognized. If the system recognizes a "Directory" Command, control is passed to the Directory Dialing Routine of FIG. **8**.

If the result of the test at step **120** is negative, a test is made at step **122** to determine if the user has spoken a "Quit" command. If so, the routine terminates. If the result of the test at step **122** is negative, the Speech Recognition System responds with a error message in step **124**. A test is then performed at step **126** to determine if a predetermined maximum error count has been reached. If not, control is transferred back to step **114**. If the predetermined maximum error count has been reached, the call and the MIN is transferred to an operator at step **128** and the routine ends.

Referring now to FIG. **6**, the Dial Routine is described in detail. At step **130**, the Speech Recognition System prompts the user with a message, such as "Phone Number Please," and applies the speaker-independent recognizer to collect the digits. A test is performed at step **132** to determine whether a digit has been collected. If not, a test is made at step **134** to determine if a predetermined timeout has elapsed. A negative response to the timeout step returns control back to step **132**. If the outcome of the test at step **134** is positive, the routine performs a test at step **136** to determine if at least the first digit of the phone number to be called has been entered. If so, another test is performed at step **138** to determine if a maximum error count has been reached. If the outcome of the test at step **138** is negative, the Speech Recognition System prompts the user to "please enter next digit" at step **140** and control returns to step **132**.

If, however, the outcome of the test at step **136** is negative, the routine tests to determine whether a maximum error count has been reached at step **142**. If not, control returns to step **130** to request the user to enter the phone number. If the outcome of the test at either step **138** or **142**

is positive, the call (with the MIN) is transferred to an operator at step **144** and the routine ends.

If the outcome of the test at step **132** is positive, the recognizer preferably responds with a short beep or other audible indication at step **146**. A test is then made at step **148** to determine if the digit collected is the last digit expected in the string. If not, the digit is saved in a string buffer at step **149** and the routine returns to step **132** to collect another digit. If, however, the outcome of step **148** is positive, all digits have been collected and the speaker-independent recognition subsystem is engaged at step **150**.

At step **152**, the subsystem attempts to verify the called number. If the verification command is not recognized, the Speech Recognition System responds at step **154** with a message such as "Error, Please repeat," and control returns back to step **152**. If the outcome of the test at step **152** is positive, the telephone number is repeated to the user at step **156** and the speaker-independent recognition subsystem is engaged.

A test is then made at step **158** to determine whether the user (by spoken command) desires to "Store" the number. If yes, the control is transferred to the Store routine of FIG. **9**. If the outcome of the test at step **158** is negative, a test **162** is made to determine whether the user has spoken (and the recognizer has recognized) a "Send" command. If not, the Speech Recognition System again plays an error message at step **164** and control is returned to step **158**. A positive response to the "Send" test made at step **162** transfers the MIN and the string dialed to the switch for outdialing. Such transfer occurs at step **166**. At step **168**, the switch dials the telephone number and connects the user to the dialed number.

The Call routine is shown in detail in FIG. **7**. The routine begins at step **170** during which the Speech Recognition System responds to the user's "Call" command with a message "Calling." At step **172**, the speaker independent recognition subsystem is engaged to obtain the user's response. A test is then made at step **174** to determine if the user speaks a "Memory" command, indicating that the number to be called is to be stored and made available for speed-dialing once a preferably two digit memory location number is subsequently received. If the outcome of the test at step **174** is positive, the Speech Recognition System queries the user "Which memory" and engages the recognition subsystem at step **176**. At step **178**, a test is made to determine if a two digit memory location has been identified by the user. If a two digit memory location has not been identified, a timeout test is performed at step **180**. If timeout occurs, the Speech Recognition System prompts the user to "Please enter next digit" at step **182** and control is returned to step **178**. A negative outcome of the timeout test also returns control back to step **178**.

A positive outcome of the test at step **178** indicates that the two digit memory code has been received. The routine then recalls the previously-stored telephone number associated with the memory code at step **184**. The Speech Recognition System then notifies issues a "Calling" message to the user at step **186**.

A user may also retrieve the stored telephone number by speaking one of the speaker-independent key words. Thus, if the outcome of the test at step **174** is negative, the routine continues at step **187** with the user speaking a valid key word associated with one or more previously-stored telephone numbers. Without limitation, such key words include HOME, OFFICE, TIME, SECRETARY, FRIEND, WORK and INFORMATION. A test is then made at step **188** to determine if the work spoken is a valid key word. If not, the system responds with an error message at step **190** and asks the user to speak the word again at step **187**. A positive outcome to the test at step **188** transfers control to step **186**.

At step **192**, the speaker-independent subsystem is again engaged to obtain a user command. A test is made at step **194** to determine whether the user desires to "Verify" the number retrieved. If a "Verify" command is spoken and recognized, the stored number associated with the key word is repeated to the user and the recognizer is engaged at step **196**. A test is then made at step **198** to determine whether the user desires to "Clear" the number retrieved and start again. If so, the Speech Recognition System responds with a "Ready" message at step **200** and control returns to step **187**.

If, however, the outcome of the test at either step **194** or step **198** is negative, a test is made at step **202** to determine if a "Send" command has been spoken and recognized. If not, the system responds with an error message at step **204** and returns to step **192**. If the user speaks the "Send" message, the Speech Recognition System responds at step **206** with a "Dialing" message. At step **208**, the System transfers the MIN and the telephone number to be dialed to the switch for outdialing. The switch dials the telephone number and connects the user to the dialed number at step **210**.

Referring now to FIG. **8**, a flowchart is shown of the Directory Dialing routine. This routine is called when the user desires to recall and dial some previously-stored telephone numbers using a previously-stored speaker-dependent name. In response to the "Directory" command from the user, the routine plays a message at step **212** to request the name in the directory. Step **212** also engages the speaker-dependent recognition subsystem. A test is then made at step **214** to determine if the name has been collected from the speaker. If no response is obtained, a timeout test is performed at step **216**. Failure of the speaker to respond before the end of the timeout causes the issuance of an error message at step **218** and the routine returns to step **214**. If the timeout occurs without the speaker's response, control is transferred back to step **212**.

A positive outcome of the test at step **214** means that the directory name spoken by the user has been recognized. The routine then continues at step **220** and recalls from the data storage subsystem a number associated with such name. At step **222**, the stored number is played to the user and the Speech Recognition System issues a "Correct" prompt. A test is then made at step **224** to determine if the number is correct. If not, control is returned back to step **212**. If the outcome of the test at step **224** is positive, then the Speech Recognition System responds with a "Dialing" message at step **226**. At step **228**, the System transfers the MIN and the telephone number to be dialed to the switch for outdialing. The switch dials the telephone number and connects the user to the dialed number at step **230**.

Referring now to the Store routine FIG. **9**, a flowchart is shown of a routine used by the subscriber to store telephone numbers for the two digit memory codes, keywords, and directory names to be later dialed. The routine begins in response to receipt of a speaker-independent "Store" command spoken by the user to update the speed-dial list whenever necessary. At step **232**, the routine responds to the command with a "Storing" message and engages the speaker-independent recognition subsystem. A test is then made at step **234** to determine whether the user desires to store two digit memory codes. If so, control is transferred to the Memory routine of FIG. **10**. If the outcome of the test at

step **234** is negative, a test is made at step **236** to determine whether the user desires to store a Directory name or a number associated with a Directory name. If so, control is transferred to the Directory Storage routine of FIG. **11**. If the outcome of the test at step **236** is negative, then the system will expect to receive a key word for association with the number to be stored at step **238**. A test is thus made at step **240** to determine whether the subscriber has spoken a valid key word. If not, the system responds with an error message at step **242** and control is returned to step **234**. If the outcome of the test at step **240** is positive, the Speech Recognition System makes an inquiry "Storing (key word), correct" and engages the speaker-independent recognition subsystem at step **244**. A test is then made at step **246** to determine if the user selection is correct. If not, the Speech Recognition System issues a "Location, please" prompt at step **248** and returns to step **240**. A positive outcome of the test at step **246** causes the system to issue a "Storing" prompt while a number associated with the keyword is stored in the data storage subsystem in step **249**. Control is then returned to step **112** in FIG. **5**.

The Memory routine is shown in FIG. **10**. The routine begins at step **250** by inquiring "Which memory" and engages the speaker-independent recognition subsystem. At step **252**, a test is made to determine if a two digit memory location number has been collected from the user. A timeout test is then performed at step **254**. If timeout occurs, the Speech Recognition System prompts the user to "Please enter next digit" at step **256** and control is returned to step **252**. A negative outcome of the timeout test also returns control back to step **252**.

A positive outcome of the test at step **252** indicates that the two digit memory code has been collected from the subscriber. At step **258**, the Speech Recognition System responds with a message "Storing memory (2 digits), correct" and engages the speaker-independent recognition subsystem. A test is then made at step **260** to confirm that the subscriber is storing the telephone number in the appropriate location. If not, the Speech Recognition System issues a "Location, please" prompt at step **262** and returns to step **252**. A positive outcome of the test at step **260** then causes the routine to test at step **264** whether the memory is already filled up with stored codes. If not, the system issues a "Storing" message and stores the number in the data storage subsystem at step **265**. If, however, the response to the test at step **264** is affirmative, the Speech Recognition System issues a "Memory full, erase?" message and engages the speaker-independent recognition subsystem at step **266**. A test is then performed at step **268** to determine whether the speaker wishes to erase a memory location. If not, control returns to step **112** in FIG. **5**.

If the subscriber desires to erase a given memory location code, the system issues a "Confirm erase" message at step **270** and engages the speaker-independent recognizer. An erase test is then performed again at step **272** to confirm that the subscriber wishes to erase the specified memory location. A negative outcome of the test at step **272** returns control to step **112** in FIG. **5**. A positive outcome of the erase test at step **272** causes the system to issue an acknowledgment at step **274**. Control then returns to step **112** in FIG. **5**.

Referring now to FIG. **11**, a flowchart is shown of the Directory Storage routine. The routine begins at step **276** in response to a positive response to the Directory inquiry performed in step **236** in FIG. **9** (i.e., receipt of the Directory command). Step **276** issues a prompt to the subscriber to determine if a "Number in Memory" has been previously associated with a Directory name. A test is then performed at step **278** to determine if the Number in memory has been previously stored. If not, the Speech Recognition System responds by issuing an "Enroll" message and engages the speaker-independent recognition subsystem at step **280**. A test is then made at step **282** to determine whether the subscriber wishes to enroll. If not, the routine terminates and returns to step **112** in FIG. **5**.

If the outcome of the test at step **278** is affirmative, the Speech Recognition System issues a message "Re-enroll" and engages the speaker-independent recognition subsystem at step **284**. A test is then made at step **286** to determine whether the subscriber wishes to re-enroll the directory name. If yes, or if the outcome of the test at step **282** is positive, the routine continues at step **288** and issues a "Name Please" message and engages the speaker-dependent recognition subsystem. The user then speaks the name a predetermined number (e.g., three (3) times) at step **290** to program the speaker-dependent subsystem. The new directory name and the associated telephone number are then stored in the data storage subsystem in step **292**. Control is then returned to step **112** in FIG. **5**.

If the outcome of the test at step **286** is negative, the Speech Recognition System issues a message "Erase" and engages the speaker-independent recognition subsystem at step **296**. A test is then made at step **298** to determine whether the subscriber wishes to erase the directory number. If so, the associated number is erased at step **300**. Control is then returned to step **112** in FIG. **5**.

It should be appreciated that the specific names of the commands are merely exemplary and should not be taken by way of limitation. Other suitable command names are of course suitable.

As noted above, each cellular telephone is uniquely identified with its MIN and ESN. In the prior art, the service provider operates the MTX **12** and keeps a database of all MIN's (and their associated ESN's). Each time a call is placed by the cellular telephone, the service provider verifies whether the MIN and the associated ESN are authorized. If the MIN and ESN of a cellular phone are not recognized in the switch database, the provider normally allows the call to proceed at least one time. If billing authorization cannot later be verified, however, the MIN and serial number are then placed on an exception list. Subsequent attempts to use the MIN will then be rejected.

Such fraudulent use of the cellular network is ameliorated by the present invention. According to another feature of the invention, the Speech Recognition System also includes a positive identification function that enables the system under certain circumstances to test whether the user associated with a received MIN (whether in or out of the service area) is an authorized subscriber to the service. For example, after a call is voice-dialed and the Send command is spoken, the system might prompt the caller to answer one or more personal questions. Thus, the caller might be requested to speak the social security and/or account number of the subscriber (purportedly associated with the MIN and ESN). Alternatively, the caller can be prompted to enter such identifying information manually through the telephone keypad. Of course, the nature and scope of the personal information requested by the system depends entirely on the system operator and the degree of security sought by the subscriber and operator.

Such a system could also be used to prevent a local call being placed from a stolen vehicle, for example, i.e., even if the MIN was recognized in the service provider's database. It could also be used to limit access to the voice-dialing

function from the subscriber's phone to certain persons. Thus the subscriber may limit his or her children or others from using the phone without the subscriber's knowledge. If the user cannot provide proper answers to such questions, the system rejects the SEND command and the call is terminated. Correct entry of the requested information enables the caller to continue his or her access to the service.

Alternatively the identification function may be implemented in conjunction with a speaker verification function in which the user's identity must first be verified with some spoken predetermined personal identification code. If the verification function is not "convinced" that the person is who he or she claims to be, the identification function is then executed to ask the follow-up questions. One such system is described in copending application Ser. No. 07/523,486, filed May 15, 1990, to Hunt et al, titled "Simultaneous Speaker-Independent Voice Recognition And Verification Over A Telephone Network," assigned to the assignee of the present invention.

The present invention has numerous advantages over the prior art. The system combines the use of both speaker-dependent and speaker-independent speech recognition in an mobile or portable telephone communications network environment. Multiple language prompts are spoken from and available simultaneously on multiple ports from a single automated telecommunications-based system. The language selected is based on the language spoken by the user. The system advantageously stores user specific speaker-dependent and speaker-independent speech information with particular user-based addresses.

The invention successfully implements speech recognition in the cellular telephone or personal communications network. Non-wireline networks provide a special challenge to both the recognition algorithm developers as well as the applications developers. The recognition algorithm in conjunction with the system application is insensitive to the radio fading, speech clipping, and speech compression conditions that occur in a non-wireline network. In addition, the recognition algorithm accommodates conditions found in the standard switched network. The invention provides a means of accurately recognizing speech that has limited distortion due to clipping or fading and provides a means of reprompting the user for input when the speech has become too distorted for accurate recognition.

Previously, only the best examples of spoken words have been used as tokens for developing speech vocabularies. By collecting speech that has been compressed or that although distorted by radio fading or clipping is still intelligible and by adding this collected speech to the speech training database, the vocabulary based on such data becomes more robust and less sensitive to these conditions. Adding the distorted but intelligible data to the training database of excellent example words allows for a more diverse statistical representation of each vocabulary word. Words that might have been previously rejected, because part of the word was clipped can now be recognized if enough intelligible information is available. If the statistical representation of the word indicates that not enough information is available for accurate recognition, the recognition system will reject the word and reprompt the user for input.

The invention also enables real-time vocabulary uploading. Previously, speech recognition vocabularies were stored in memory and were not updated or totally changed during the operation of an application. The present invention solves the problem of allowing speech recognition vocabularies to be loaded into a recognition system while an application is in operation. This capability allows for less high speed memory to be used in a system, because an entire set of vocabularies is not required to be resident in memory at one time. This capability also allows for efficient memory management when multiple languages are used, because only the language required by the system user need be resident in memory at any one time. Previously, all possible languages required by users would have been stored in memory simultaneously.

Both continuous and discrete recognition techniques are used in the Speech Recognition System of the present invention.

It should be appreciated by those skilled in the art that the specific embodiments disclosed above may be readily utilized as a basis for modifying or designing other structures for carrying out the same purposes of the present invention. It should also be realized by those skilled in the art that such equivalent constructions do not depart from the spirit and scope of the invention as set forth in the appended claims.

What is claimed is:

1. A speech recognition method for a mobile telecommunication system which includes a voice recognizer capable of recognizing commands and characters received from a mobile telecommunication user, the method comprising the steps of:

receiving a command from the mobile telecommunication user;

determining whether the command is a first or second command type;

if the command is the first command type, collecting digits representing a telephone number to be dialed received from the mobile telecommunication user; and

if the command is the second command type, determining whether a previously stored telephone number is associated with a keyword received from the mobile telecommunication user.

2. The method according to claim **1**, wherein the keyword is a name.

3. The method according to claim **1**, wherein the keyword is a location or relationship modifier.

4. The method according to claim **3** wherein the location modifier is home, work or office.

5. The method according to claim **1**, further comprising the steps of verifying the command and initiating a telecommunication call with the mobile telecommunication system.

6. The method according to claim **1**, further comprising the step of prompting the mobile telecommunication user to enter information needed for the first or the second command type.

7. The method according to claim **1**, further comprising the step of verifying the mobile telecommunication user's identity.

8. The apparatus according to claim **7** wherein the speech recognition apparatus is connected to the mobile telecommunication switch as an external peripheral.

9. The apparatus according to claim **7** wherein the memory is connected to the controller through a data network.

10. The apparatus according to claim **7** wherein the speech recognition apparatus is a shared resource that can be accessed by more than one mobile telecommunication switch.

11. The apparatus according to claim **7** wherein the speech recognition apparatus is connected to a non-mobile telecommunication switch and the interface communicates with the

mobile telecommunication switch through the non-mobile telecommunication switch.

12. The apparatus according to claim **7** wherein the voice recognizer is also capable of recognizing commands and characters received through the interface from a non-mobile telecommunication user.

13. The apparatus according to claim **7** wherein the mobile telecommunication system interfaces with a communication network.

14. The apparatus according to claim **13** wherein the communication network is a circuit switched network.

15. The apparatus according to claim **13** wherein the communication network is a packet switched network.

16. The method according to claim **1** wherein the keyword is associated with a type of data requested by the mobile telecommunication user.

17. The method according to claim **16** wherein the keyword includes time and information.

18. The method according to claim **16** further comprising the step of having the mobile telecommunication user subscribe to a service that includes a predetermined list of keywords.

19. The method according to claim **1** further comprising the step of having the mobile telecommunication system establish a path connection between the mobile telecommunication user and the voice recognizer.

20. A speech recognition apparatus for a mobile telecommunication system, the apparatus comprising:

a memory;

an interface with a mobile telecommunication switch;

a voice recognizer capable of recognizing commands and characters received through the interface from a mobile telecommunication user; and

a controller, coupled to the memory and the voice recognizer, arranged to determine whether a speech input from the user is a call command, if the command is a first call command type, collect digits representing a telephone number to be dialed spoken by the user, if the command is a second call command type, determine whether a previously stored telephone number is associated with a reference code received from the mobile telecommunication user.

21. The apparatus according to claim **20**, wherein the reference code is a name.

22. The apparatus according to claim **20**, wherein the reference code is a location or relationship modifier.

23. The apparatus according to claim **20** further comprising means for verifying the mobile telecommunication user's identity.

24. The apparatus according to claim **20** wherein the interface communicates with the mobile telecommunication switch using at least one trunk.

25. The apparatus according to claim **24** wherein the at least one trunk is a digital trunk.

26. A voice activated dialing system for a wireless communication user, the system comprising:

at least one wireless telecommunication switch;

an interface with the at least one wireless telecommunication switch;

a voice recognizer capable of recognizing commands and characters received through the interface from the wireless communication user; and

a controller, coupled to the voice recognizer, arranged to determine whether a speech input from the wireless communication user is a call command, if the command is a first call command type, collect digits representing a telephone number to be dialed spoken by the user, if the command is a second call command type, determine whether a previously stored telephone number is associated with a reference code received from the wireless communication user.

27. The system according to claim **26**, wherein the reference code is a name.

28. The system according to claim **26** wherein the interface communicates with the at least one wireless telecommunication switch using at least one trunk.

29. The system according to claim **26** wherein the voice recognizer and the controller form a speech recognition node that is communicatively coupleable to the at least one wireless telecommunication switch as an external peripheral.

30. The system according to claim **26** wherein the voice recognizer and the controller form a speech recognition node that acts like a shared resource that can be accessed by more than one telecommunication switch.

* * * * *

EXHIBIT 2

125 SUMMER STREET BOSTON MA 02110-1618
T 617 443 9292 F 617 443 0004 WWW.BROMSUN.COM

BROMBERG & SUNSTEIN LLP

ERIK PAUL BELT
T 617 443 9292 x260
EBELT@BROMSUN.COM

April 25, 2005

VIA FACSIMILE AND FIRST CLASS MAIL

Robert S. Frank, Jr., Esq.
Choate, Hall & Stewart LLP
Exchange Place
53 State Street
Boston, MA 02109-2809

Re ScanSoft, Inc. v. Voice Signal Technologies, Inc., et al.
Civil Action No. 04-10353-PBS
Our File: 02639/00509

Dear Bob:

The parties have agreed to exchange proposed claim constructions for terms of ScanSoft's '966 patent and of VST"s '630 patent previously identified in the parties' letters of March 30, 2005. ScanSoft's proposed claim constructions for these terms appear below.

1. Terms of '966 Patent. In your letter of March 30th, you identified six terms or phrases of ScanSoft's '966 patent that VST believes will require construction by the Court. ScanSoft disagrees with VST's contention that all such terms and phrases will require construction. For the purpose of this exchange, however, ScanSoft provides the following constructions. In each case, the meaning of the term is as it would have been understood by one of ordinary skill in the art.

[redacted]	[redacted]
"A speech recognition method . . ."	A process for recognizing commands and characters spoken by a user
" . . . a mobile telecommunications system which includes . . ."	"mobile telecommunications system" refers to "cellular, satelite [*sic*] and personal communications network environments." Col. 3, *ll.* 46-48. A network environment includes all components of a network, from the central office switching equipment, if any, to the

	mobile units (e.g., telephones, pdas, and other communications devices) used therein.
" . . . a voice recognizer"	A "voice recognizer" comprises software and/or hardware used to process sounds spoken by a user.
". . . comprising the steps of"	"Comprising" means including but not limited to the recited steps.
" . . . determining whether the command is a first or second command type"	A process performed by the speech recognition system to recognize whether the user has spoken a first or second command type so that the system can then perform the routines associated with the first or second command type.
" . . . a first command type"	A "command" is a word, phrase, or numeric digit, alone or in combination with other words, phrases, and/or digits, used to direct the speech recognition system to take some action. A "first command type" is one category of commands recognized by the speech recognition system.
" . . . a second command type"	A "command" is a word, phrase, or numeric digit, alone or in combination with other words, phrases, and/or digits, used to direct the speech recognition system to take some action. A "second command type" is another category of commands recognized by the speech recognition system.
". . . collecting digits representing a telephone number . . ."	The speech recognition system collects a series of numbers that the system recognizes as a complete telephone number.
". . . determining whether a previously-stored telephone number is associated with a keyword . . ."	The speech recognition system checks if there is a match between the keyword spoken by the user and a telephone number previously stored in the system and previously associated with a keyword. A "keyword" is any word, phrase, or numeric digit, or combination thereof, used to recall a previously-stored telephone number.

comparing said command word portion and said pause portion of said at least one received audio command with said command word portion and said pause portion, respectively, of said speech recognition data;	An input command having a word portion of at least one syllable in length, ~200 milliseconds, and a pause portion of at least one syllable in length, ~200 milliseconds, is compared to speech recognition model data having a word portion at least one syllable in length, ~200 milliseconds, and a pause portion at least one syllable in length, ~200 milliseconds,.
generating at least one control signal based on said comparison;	A "control signal" is a signal that directs or manages the operation of the electrical device that is controlled by the control signal.
controlling power delivered to an electrical device in response to said at least one control signal for operating the electrical device in response to said at least one received audio command;	Electrical power is turned on or off, or increased or decreased, to an electrical device in response to a control signal generated in response to an input audio command.
analyzing the pause portion of the received audio command for spectral content; and	"Content" refers to what is contained within something. In this case, the content of the pause portion of the received audio command must be "spectral." A speech signal contains spectral components in the sense of energy at various different specific frequencies. Thus, "spectral content" means that the pause portion of the received audio command must have energy at various different frequencies.
preventing operation of the electrical device when the spectral content is dynamic.	"Dynamic" refers to something that is in motion or is changing. Thus, this claim element means that operation of the electrical device is prevented when the at least one syllable long pause portion of the input audio command has energy at various different frequencies and which is moving or changing.

14. The method of claim 7, and further comprising:	
receiving background noise data in conjunction with receiving said at least one audio command;	"Background noise" is part of the user's acoustic environment that includes everything besides the spoken user input. Thus, an acoustic signal input is received that includes a spoken command from the user, and everything else in the background acoustic environment.
ascertaining a first energy content for the command word portion of the received audio command;	"Content" refers to what is contained within something. In this case, the content of the command word portion of the received audio command must be 'energy' content. The energy of a speech signal refers to its force or strength and in visual representations is characterized by the amplitude or size of the speech signal. Thus, "first energy content" means that the command word portion of the received audio command is characterized by a first kind of inherent force or strength within it.
ascertaining a second energy content for the received background noise data;	The system must also determine a "second energy content," which is characteristic of the inherent force or strength contained within the background noise signal.
comparing the first and second energy contents and generating a corresponding energy comparison value; and	A signal is generated that has a characteristic value representative of the correspondence between the first and second energy contents.
preventing the generation of said at least one control signal when said energy comparison value is below a predetermined level.	The control signal cannot be created if the energy comparison value is less than some pre-set value.

16. A method of activating an electrical device ...	A process for turning on an electrical device.
... through at least one audio command from a user, the method comprising the steps of:	An audio command is a word, phrase, or numeric digit, alone or in combination with other words, phrases, and/or digits which is spoken by a user to turn on the electrical device.
recording speech recognition data having a command word portion and a pause portion, each of the speech recognition data portions being at least one syllable in length;	The "speech recognition data" is the speech model information that unknown input speech is compared to. A pause refers to a momentary stop in an ongoing event. The patent and its prosecution history define a syllable as a length of time at least about 200 milliseconds long. A "pause portion" is that part of the speech recognition data comprising a pause. Thus, this claim element requires that the speech model information includes a momentary stop in speech that is at least about 200 milliseconds long.
receiving at least one audio command from a user, the at least one audio command having first and second command word portions and a first, second and third pause portions, each of the audio command portions being at least one syllable in length, said second pause portion having one syllable in duration before said first command word portion and said third pause portion having one syllable in duration after said second command word portion;	The system gets from a user an audio command that has five different parts, all of which are at least 200 milliseconds long. Two word parts are separated by one syllable part. And two other different syllable parts respectively precede and follow the word parts.
comparing said command word portion and said pause portion of said at least one received audio command with said command word portion and said pause portion, respectively, of said speech recognition data;	*This claim element is indefinite and fails to satisfy the requirements of 35 U.S.C. §112. Therefore, it cannot be correctly construed.*
generating at least one control signal based on said comparison;	A control signal is a signal that directs or manages the operation of the electrical device that is controlled by the control signal.

controlling power delivered to an electrical device in response to said at least one control signal for operating the electrical device in response to said at least one received audio command.	Electrical power is turned on or off, or increased or decreased, to an electrical device in response to the control signal generated in response to a five part input audio command of the type required above.

Very truly yours,

Erik Paul Belt

EPB/id

02639/00509 381236.1

EXHIBIT 3

CHOATE, HALL & STEWART LLP
EXCHANGE PLACE
53 STATE STREET
BOSTON, MASSACHUSETTS 02109-2804
T (617) 248-5000 F (617) 248-4000
www.choate.com

ROBERT S. FRANK, JR.
DIRECT DIAL: (617) 248-5207
EMAIL: RFRANK@CHOATE.COM

April 25, 2005

VIA FACSIMILE and
FIRST CLASS MAIL

Lee Carl Bromberg, Esq.
Bromberg & Sunstein LLP
125 Summer Street
Boston, MA 02110-1618

Re: ***ScanSoft, Inc. v. Voice Signal Technologies, Inc., et al.,***
Civil Action No. 04-10353 PBS
Your File 2639/509

Dear Lee:

The following are Voice Signal's proposed constructions of the terms and phrases in the '966 patent which Voice Signal believes require construction. You have not identified any term which ScanSoft believes requires construction.

1. A speech recognition method *for a mobile telecommunication system* which includes a voice recognizer . . .

   - A speech recognition method that is used by a mobile telecommunication system, as distinguished from a method used in a mobile (*e.g.*, cellular) telephone. A mobile telecommunication system is a network that connects mobile telecommunications customers each having a mobile telephone to other telephone customers. The mobile telecommunication system must have a voice recognizer that recognizes spoken commands that are received by it over a cellular or other non-wireline network.

2. A speech recognition method for a mobile telecommunications system which includes a voice recognizer . . . comprising the steps of . . . receiving a *command* from the mobile telecommunication user;

   - The mobile telecommunication system, which includes a voice recognizer, must receive sounds spoken by a mobile telecommunications user that it recognizes as one of a pre-established group of command words.

3. A speech recognition method for a mobile telecommunications system which includes a voice recognizer . . . comprising the steps of . . . *determining whether the command* is a *first* [command type] or *second command type*

- The mobile telecommunication system, which includes a voice recognizer, must determine whether the command word that it has recognized indicates that the user will next speak a series of numbers or a keyword.

4. A speech recognition method for a mobile telecommunications system...comprising the steps of . . . *collecting digits representing a phone number* to be dialed received from the mobile telecommunication user

- The mobile telecommunications system, which includes a voice recognizer, must collect a series of spoken numbers that the user wishes to dial.

5. A speech recognition method for a mobile communications system . . . comprising the steps of . . . *determining whether a previously-stored telephone number is associated with a keyword*

- The mobile telecommunications system must determine whether a keyword spoken by a user is associated in the memory (the data storage subsystem) of its speech recognition system with a telephone number.

Set forth below are Voice Signal's proposed constructions of the terms in the '630 patent that require construction:

A. A pause portion . . . being at least one syllable in length

- A pause portion, the duration of which is within the known duration range for one syllable of speech

B. "spectral content"

- Sounds are made up of energy at different frequencies. Spectral content is the amount of energy in each frequency that makes up a sound or a series of sounds over a specified time.

"energy content"

- Energy content is the total amount of energy detected in a given segment of sound (which could be a command word portion, background noise, or a pause).

". . . preventing the operation of the electrical device when the spectral content is dynamic"

- When the spectral content is changing more than background noise would be expected to change, the spectral content is said to be dynamic. If the spectral content of the pause portion of the received audio command is dynamic, the electrical device is prevented from operating.

Voice Signal reserves the right to vary the phrasing (as distinguished from the substance) of these definitions in its claim construction submission to the Court.

Sincerely yours,

Robert S. Frank, Jr.

RSFJr/kf

CHOATE
HALL &
STEWART

Fax Transmittal Sheet

RECIPIENT	COMPANY	FAX	PHONE
Lee C. Bromberg, Esq.	Bromberg & Sunstein LLP	617-443-0004	617-443-9292

FROM	Robert S. Frank, Jr.	NUMBER OF PAGES	4
DATE	April 25, 2005	CLIENT NUMBER	2004207-0003
PHONE	(617) 248-5207	OPERATOR	TIME SENT

COMMENTS **Please see attached correspondence.**

RETURN BY Inter-office Mail Hold for pick-up